Ab initio Simulations of the P-Cluster in Nitrogenase and Multi-Scale Methods

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften vorgelegt von

Sascha Michael Hemmen
aus Saarbrücken

genehmigt von der Fakultät für Natur- und Materialwissenschaften
der Technischen Universität Clausthal.

Tag der mündlichen Prüfung: Freitag, 18. April 2008

Ab initio Simulations of the P-Cluster in Nitrogenase and Multi-Scale Methods

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
vorgelegt von

Sascha Michael Hemmen
aus Saarbrücken

genehmigt von der Fakultät für Natur- und Materialwissenschaften
der Technischen Universität Clausthal

Tag der mündlichen Prüfung: 18. April 2008

Hiermit erkläre ich an Eides Statt, dass ich die bei der Fakultät für Material- und Naturwissenschaften der Technischen Universität Clausthal eingereichte Dissertation selbständig und ohne unerlaubte Hilfe verfasst und die benutzten Hilfsmittel vollständig angegeben habe.

Clausthal, 11. März 2008 Sascha Hemmen

Die Arbeit wurde angefertigt am
Institut für Theoretische Physik der Technischen Universität Clausthal.

Vorsitzender der Promotionskommission: Prof. Dr. Karl-Heinz Spitzer

Hauptberichterstatter: Prof. Dr. Peter E. Blöchl

Berichterstatter: Prof. Dr. Tom Kirchner

Contents

Abstract

Nitrogen is one of the major elements in organic and biological molecules. While nitrogen occurs in abundance in the atmosphere, the metabolism of nearly all plants and creatures cannot directly access it. Only certain bacteria, called diazotrophs, contain the enzyme Nitrogenase which is able to change atmospheric nitrogen into a metabolic useful form, ammonia.

The biological nitrogen fixation is driven by a multi-step proton and electron transfer process. The electrons needed by the active site are provided by the P-cluster. The P-cluster is an iron-sulfur cluster (Fe_8S_7) with a complex structure. It is known that it changes its structure under change of the oxidation state. In this thesis, I have investigated the mechanism of these structural changes and the different spin structures of the P-cluster compared to the results of experimental data. The P-cluster has been simulated in vacuum with only the next neighbors of the protein environment.

The number of link bonds of the P-cluster to the protein environment varies with the oxidation state. It is important whether the P-cluster can control the proton and electron transfer through the enzyme. Breaking and formation of bonds requires a quantum mechanical simulation of the cluster. QM simulations are limited to several hundred atoms in the system. For the changes of the P-cluster do not modify the chemistry of the rest of the protein, molecular mechanical methods can be employed to simulate the environment. The combination of quantum mechanical and molecular mechanical simulations is called QM/MM. The QM/MM method gives access to different length scales in the simulation, implying different time scales, too. Oversampling of the environment is important to let the protein respond to the structural changes in the reaction center.

Part of this work has been to improve the preliminary implementation of the QM/MM method in the CP-PAW program package. Now, larger systems can be treated and the user interface has been improved, which enables the use of structure files from the protein data base. The AMBER force field is well known for its accuracy when treating biological systems and has been implemented.

Including the protein environment into the simulation forces the inclusion of a solvent, too. The TIP3P water model has been employed to describe water within the AMBER force field. This method requires a large number of additional atoms in the simulation cell. Another solvation model, the conductor-like screening model (COSMO), has also been tested. Charges placed on the surface of a cavity around the solute simulate the solvent's effects. A strategy to determine the solvation radii for the atom types in COSMO has been developed.

1. Introduction

Today's computer simulations have become a powerful tool in science. Steadily increasing computer power and the improvement of algorithms make it possible to simulate more and more complex systems on the computer. In the present thesis I will show the combination of different simulation methods to get a better understanding of biological nitrogen fixation by Nitrogenase. But, what is the reason to focus on Nitrogenase?

One of the major components of organisms, besides carbon, hydrogen and oxygen, is nitrogen. Nearly 10% of the biological dry mass is nitrogen. Although up to 80% of our atmosphere consists of dinitrogen, this nitrogen is metabolic inaccessible to most living creatures and plants. The triple bond of dinitrogen determines its inert behavior. To separate the nitrogen atoms all three bonds have to be broken which costs a large amount of energy.

How does nitrogen come into the biological cycle? The answer is nitrogen fixation. On the one hand, there is the un-biological nitrogen fixation. First, lightning can provide enough energy to break the triple bond. The results are simple nitrogen compounds, as ammonia salts, simple amino acids or cyan salts. Urey and Miller [1] demonstrated in their experiments that a model of an early earth atmosphere[1] yields simple amino acids, ammonia and simple carbohydrates. Nevertheless, lightning alone cannot provide enough nitrogen for the biological cycle. Another way to bring nitrogen into this cycle is the combustion of organisms. However, new material is not obtained in this case. The third method of un-biological nitrogen fixation is the Haber-Bosch cycle. This is a method used by industry which requires high temperature and high pressure to break the triple bond of dinitrogen.

On the other hand, nature uses the biological nitrogen fixation. Special bacteria employ Nitrogenase for this purpose. These bacteria can be found living in symbiosis with plants like peas, clover or beans. Nitrogenase is able to break the triple bond of dinitrogen under ambient conditions. Although Nitrogenase is one of the oldest enzymes in evolution's history, its structure has been first revealed in 1992. Since then, the reaction mechanism of the biological nitrogen fixation has been focused in the work of many scientists. The complete catalytic cycle has been proposed by Kästner in his PhD thesis [2] 2004. Today, there are still open questions concerning the mechanism.

The nitrogen fixation needs electrons and protons at the reaction center. The reaction center is one of three iron-sulfur clusters in the protein. The other iron-sulfur clusters take part in the electron transport mechanism. One of these clusters, the P-cluster, is of special interest. Experiments have shown that the P-cluster changes its structure when it changes its oxidation state. The first

[1] This has been a mixture of water, methane, carbon dioxide and ammonia. Later ammonia was substituted by dinitrogen which did not change the results.

part of my work on this thesis has been to understand the mechanism of this structural change. Starting from experimental data I simulated the P-cluster in its different oxidation states from first principles. This is reported in Chapter 6. Details of the methods used in the calculations are reported in Chapter 4. I also give a detailed view of the spin distribution of the cluster and its influence of the cluster's structure.

The structural changes of the P-cluster influence the nearby protein environment. Experiments have shown that the cluster builds two new links to the protein environment in its oxidized state. While the controlling of the proton transport mechanism is still unknown, these changes of the environment can give the right hints to understand the mechanism. However, density functional (DFT) calculations cannot handle the amount of atoms which are needed to simulate the changes in the environment. Other simulation techniques have to be employed. I used a combined quantum mechanical and molecular mechanical (QM/MM) method as basic approach for the multi-scale simulations. Multi-scale simulations allow simulations on different length and time scales. The preliminary QM/MM implementation in the CP-PAW program package has been extended during the work on this thesis. The AMBER force field has been implemented for it is most accurate when describing proteins. Details of the QM/MM implementation are presented in Chapter 2. Chapter 3 deals with the AMBER force field.

The second part of the thesis deals with the applications of those methods. The details of the implementation of AMBER into the CP-PAW framework are reported in Chapter 5. There, the need of a solvent is also discussed. For proteins, the solvent of choice is water. The TIP3P water model is implemented in the AMBER force field. Some applications which show the benefits of the AMBER force field in QM/MM simulations together with a water model are also presented in this chapter.

In Chapter 7 another solvation model is presented. The conductor-like screening model (COSMO) simulates the dielectric properties of a solvent. The advantage of this method is, that different solvents can be described with the same method. The user has only to change the dielectric constant. However, the calculated solvation energies depend strongly on the chosen solvation radii of the atoms. Thus, it is necessary to find the optimal solvation radii for each solvent.

Part I.

Theory and Methods

2. Combined Quantum Mechanics and Molecular Mechanics Methods

2.1. Introduction

The size of biological systems is one of the major challenges to theorists when studying chemical reactions. Typical systems are enzymes which often contain several hundreds of atoms. Such proteins - chains of amino acids - build their functional structure in a solvent, which is naturally water. Thus, the solvent cannot be ignored. The solvent increases the number of atoms in these systems to thousands. Today hybrid quantum mechanical/molecular mechanical (QM/MM) methods are the best choice for studying their chemical reactions. Molecular mechanical (MM) methods have long been used to study the structural and energetic properties of biological systems. For example, they are used successfully in the study of protein folding or protein-substrate interaction. The forces are derived from classical potentials which are constructed by force fields. The weakness of such force fields is that they are parameterized against ground state systems. Therefore, they are unable to describe the breaking or formation of chemical bonds. Enzymatical reactions (the breaking and formation of bonds) often take place in special regions of the protein. While the protein environment is also important for the reaction, the active sites are small enough to be described by quantum mechanical (QM) methods. QM methods describe the electrons explicitly. They require little or no parameterization and are very accurate. The disadvantage of these methods is that they are computationally expensive, thus limiting the system size. Hybrid methods are a way of combining the computational speed of MM methods with the high accuracy and flexibility of QM methods.

The first combination of MM and QM methods was reported by Warshel and Levitt in the mid-1970s [3]. They studied the enzymic reaction of lysozyme. In the catalytic mechanism for the cleavage of hexasaccharides by lysozyme they found an oxocarbonium ion as intermediate which is stabilized by the environment. While the energy and charge distribution of atoms which are directly involved in the reaction were evaluated quantum mechanically, the potential energy surface of the rest of the system was evaluated classically. A pure QM calculation did not lead to the same result because the QM system was not able to stabilize the transition state. Today, more than 30 years later, the subject is still developing. In the last years the rate of publications of QM/MM applications increased rapidly. One reason is the availability of implementations in commercial packages (e.g. Gaussian98 or CHARMM). In the following sections

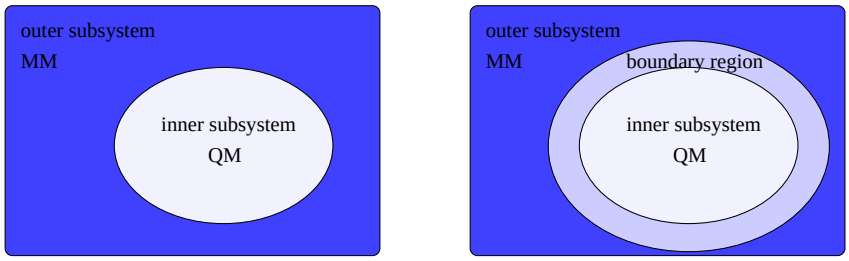
a short overview of the range of approaches of QM/MM calculations is given.

2.2. Terminology

In the first step the entire system (S) is divided into an inner system (I) and an outer system (O). The inner system is treated quantum-mechanically and the outer system is described by a classical force field. Therefore we speak of a QM and an MM region. Each atom of the entire system belongs to one of the two subsystems. It is not possible to write the total energy of the entire system as a sum of the subsystems. Because the interaction between the subsystems is sufficiently strong (e.g. by the presence of a chemical bond), the boundary has to be treated in a special way. There are two approaches to terminate the QM region:

- Link atoms. In this approach link atoms will be introduced in the boundary region. The QM subsystem will be chemically saturated with atoms. They do not belong to S, ensuring that the QM system can be treated alone without producing errors. The positions of the link atoms can either be viewed as independent variables or as a function of the positions of the atoms in the systems I and O. Figure 2.1(a) shows this scheme. The system of all link atoms will be denoted as L.

- Boundary Region. In this approach a boundary region is defined. Atoms in this region will appear in both calculations, QM and MM. Figure 2.1(b) shows this approach.



(a) Inner and outer region (b) Boundary region

Figure 2.1: Scheme of the entire system S divided in subsystems O and I with and without boundary region.

2.3. QM/MM Energy Expressions

2.3.1. Subtractive Schemes

A subtractive scheme is an entire MM system where the region of special interest is cut out. The cut subsystem is replaced by a system which is treated on the QM level. In the end three calculations are necessary in subtractive schemes:

1. a calculation of the whole system on the MM level
2. a QM calculation of the inner system
3. again a calculation of the inner system on the MM level

The total energy of the QM/MM system is obtained by summing (1) and (2) and subtracting (3) to avoid double counting.

$$E_{QM/MM}(S) = E_{MM}(S) + E_{QM}(I+L) - E_{MM}(I+L) \tag{2.3.1}$$

The subscript indicates the level of the calculation. Equation (2.3.1) is written for systems with link atoms, where the inner system I is capped by link atoms L. For schemes with special boundary atoms as described above, L refers to these atoms. The calculations on the inner subsystem are not only performed on I but on $I+L$. $I+L$ describes the capped inner system.

The advantage of this formulation is its simplicity. Because there is no need of any coupling terms, the QM and MM routines do not have to be modified. Any errors from the link atoms are subtracted out.

The calculation of the inner system on the MM level requires the parameterization of the QM part. Unfortunately, many force fields are only parameterized for special systems. As an example, the AMBER force field has been constructed for biological systems like amino acids. It has not been parameterized for metal clusters like the P-cluster in Nitrogenase. In such cases the user has to extend the force field of choice for his belongings. However, the additional parameters do not have to be very accurate because the error produced by bad parameters in the inner subsystem is subtracted out of the total energy.

The electrostatic interaction between the subsystems deserves closer attention. During a chemical reaction the charge distribution in the inner subsystem can change which cannot be described by rigid point charges properly. It is also desirable that the QM charge density is affected by the outer charge distribution. The way these problems are solved will be described in later sections.

2.3.2. Additive Schemes

Particularly in the area of biochemistry the additive schemes are of widespread use. Important examples are the AMBER [4, 5, 6] and CHARMM [7, 8, 9, 10] based implementations[1].

One can easily see that the subtractive and additive models described here are closely related. Bakowies and Thiel use the subtractive models to derive and discuss link-atom corrected additive models [11].

Additive schemes are systems where the energy of each subsystem is calculated by its own. The total QM/MM energy of the entire system is the sum of the subsystems plus a coupling term:

$$E_{QM/MM}(S) = E_{MM}(O) + E_{QM}(I+L) + E_{QM-MM}(I,O) \tag{2.3.2}$$

[1] It is important to note that one has to distinguish between methods and program packages where the methods are implemented and which have the same name as the methods they use.

The additional coupling term $E_{QM-MM}(I,O)$ contains the interaction of the two subsystems[2]. The QM/MM energy can be derived from the subtractive energy expression in equation (2.3.1). The MM terms are decomposed into terms which depend on exclusive sets of atoms:

$$E_{MM}(S) = E_{MM}(O) + E_{MM}(I) + E_{MM}(I,O) \quad (2.3.3)$$

$$E_{MM}(I+L) = E_{MM}(I) + E_{MM}(L) + E_{MM}(I,L) \quad (2.3.4)$$

Substituting these into equation (2.3.1) we get:

$$\begin{aligned} E_{QM/MM}(S) &= E_{MM}(O) + E_{QM}(I+L) + E_{MM}(I,O) \\ &\quad - [E_{MM}(L) + E_{MM}(I,L)] \end{aligned} \quad (2.3.5)$$

The energy on the MM level of the inner subsystem $E_{MM}(I)$ cancels out. For link-atom-based systems the coupling energy on the MM level between inner and outer subsystem $E_{MM}(I,O)$ can be identified with the correction term $E_{QM-MM}(I,O)$ in equation (2.3.2). The subtractive terms in equation (2.3.5) can be understood as a link atom correction. Link atoms are only part of the inner system. Their interaction among each other as well as their interaction with the other atoms of the inner system is described by $E_{QM}(I+L)$.

Any particular QM/MM method is defined by the exact form of the QM/MM coupling term $E_{QM-MM}(I,O)$ in equation (2.3.2). This term gathers the interactions between the systems namely the electrostatic interactions, the van-der-Waal interactions and the bonded interactions between the QM and the MM atoms. This leads to:

$$E_{QM/MM}(I,O) = E^{el}_{QM/MM} + E^{vdW}_{QM/MM} + E^{b}_{QM/MM} \quad (2.3.6)$$

In section 2.3.5 I will discuss the van-der-Waals and electrostatic contributions to $E_{QM-MM}(I,O)$ in more detail.

2.3.3. The choice of the QM model

Since the choice of the QM model does not fundamentally affect the design of a QM/MM scheme I will not discuss it in detail.

A large number of *ab initio* schemes based on Hartee-Fock [4, 8, 12, 13] and density functional [6, 14, 15, 16, 9] approaches have been implemented during the last years. Recently a number of approaches based on Car-Parrinello DFT codes has been reported [17, 18, 19]. In this thesis I present our own QM/MM implementation in the CP-PAW package. This will be discussed later in detail.

2.3.4. The choice of the MM model

Many force fields are parameterized for special systems. The choice of the right force field depends on the problem which shall be solved. Also the choice of the

[2] It has to be noted that there is a difference in the notation of the energy terms. $E_{QM}(I+L)$ denotes the energy of the inner system containing I and L. $E_{QM}(I,L)$ is the interaction between I and L.

MM scheme depends on whether the additive or subtractive schemes are chosen. Following the discussion above a subtractive scheme can be used with any force field. In the case of additive schemes one has to distinguish between different kinds of force fields. Different classical approaches differ in the handling of both bonded and non-bonded interactions which have an influence on the treatment of the boundary. The most important distinction is that between:

- valence force fields, like the biomolecular force fields as CHARM [20], AMBER [21], GROMOS [22] and more general purpose force fields like MM3 [23]. These force fields are constructed from the energy terms like bond stretches, angle bends, etc. and use point charge models. Force fields used for (bio)molecular systems are all valence force fields.

- ionic force fields, where the principal terms are the electrostatic and short rang forces (e.g. [24]).

MM schemes based on link atoms are easier to construct with valence force fields. The ionic class of force fields can be used in boundary-region additive schemes, but only if the force field charges can generate the correct electrostatic potential in the QM region. This approach is necessary to get the correct interionic forces and it is a good choice for highly ionic materials.

2.3.5. Non-bonded interactions between the subsystems

In many MM schemes the non-bonded interactions contain electrostatic and short-range (van-der-Waals) forces.

Van-der-Waals terms

In most cases the van-der-Waals interaction is described by a Lennard-Jones potential. Sometimes alternative functional forms are used, e.g. with an exponential repulsive term. However, the exact form of this term does not affect the discussion here. The problem is the availability and the suitability of the right parameters for the inner subsystem. It is a common problem that parameters for certain QM atoms do not exist in the force field. If parameters exist for these atoms, they belong to special atomic configurations. During a reaction the configuration of the atoms can change and therefore one has to use a different parameter set. The difficulty would be to decide whether one should switch the parameter set between 'educt' and 'product' and at which point along the reaction path the set should be switched. Finally the MM parameters are not transferable and are only valid within the parameterization they were derived for, that is, for MM-MM interactions.

In practice these problems can be neglected due to the short-range nature of the van-der-Waals interaction. Although each atom of the inner system interacts with all outer atoms, only atoms at the boundary between the subsystems are affected by bad van-der-Waals parameters. If someone is concerned whether this has influence on the outcome of the simulation he can shift the QM-MM boundary further away from the bad parameterized QM atoms.

Electrostatic terms

Bakowies and Thiel [11] defined three models for the electrostatic coupling between the QM charge density and the charge model used in the MM region. The models are classified as mechanical embedding (model A), electrostatic embedding (model B) and polarized embedding (model C). In the following a short overview of these models is given.

Mechanical Embedding. In a mechanical-embedding scheme the QM calculation is essentially performed in the gas phase, without electrostatic coupling of the environment. The electrostatic interaction of the QM and the MM part is either omitted or performed on the MM level. The charge model of the MM method - usually atomic point charges - is applied on the atoms in the QM region. Although this is a very straightforward model there are some disadvantages to mention. In this model the charges of the outer region do not interact with the QM density. Therefore the QM density is not polarized by the environment. Also the charge distribution in the QM region can change during a reaction which requires an update of the charge model. This leads to discontinuities in the potential-energy surface. The MM charge model depends on the other force field parameters. Together with these they yield a balanced description of conformational or structural preferences, rather than reproducing the true charge distribution.

Electrostatic Embedding. In this model the MM charge distribution appears as an external charge distribution in the QM Hamiltonian, which is augmented by an additional term:

$$\hat{\boldsymbol{H}}^{el}_{QM-MM} = -\sum_{i}^{electrons} \sum_{M \in O} \frac{q_M}{|\mathbf{r}_i - \mathbf{R}_M|} + \sum_{\alpha \in I+L} \sum_{M \in O} \frac{q_M Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{R}_M|} \quad (2.3.7)$$

where q_M are the MM point charges and Z_α the nuclear charges of the QM atoms; index i runs over all electrons, M over the point charges and α over the QM atoms. In such schemes the electrostatic interaction of the QM and MM part is treated on the QM level. Changes in the charge distribution of the environment affect the QM region and polarize it. Special care is required in the boundary region when MM charges come too close to the QM density. This can result in an over-polarization of the charge density. Especially in the case of link bonds one has to take care about this problem.

On the other hand the problem remains that the electrostatic parameters of the MM region are not designed to represent the real charge distribution. But experience shows that this common method yields reasonable results, at least for the combination of a QM density with one of the widely used biomolecular force fields [25]. The advantage of this method is that force fields provide all atomic charges and the inclusion in the QM Hamiltonian is efficient. The model of electrostatic embedding is the most popular scheme today.

Polarized Embedding. The next logical step is to create a model where the polarized charge density of the QM system interacts with the charges in the MM system. Thus one system is polarized by the other. This makes most sense for force fields which incorporate polarization but there is a lack of well-established

polarizable biomolecular MM force fields. For an overview of actual models I refer to the review article by Senn and Thiel [25].

These models have been implemented for the first time in a computer program called MNDO/MM by Bakowies and Thiel [11]. It uses either MNDO [26] or AM1 [27] wavefunctions and the MM3 force field [23]. The ChemShell program package [28] is another widely used package in which these models have been implemented.

2.4. Termination of the QM region

In this section I will discuss the problems that arise when covalent bonds are cut by the QM-MM boundary. The simplest solution of this problem would be to define the subsystems in a way that no bonds cross the borders. This is quite easy to realize in solvation studies where only the solute and optionally the first solvation shell are treated at the QM level and the rest of the solvent is placed in the MM part. Studying enzymatic reactions, it is unavoidable to cut through covalent bonds. Only certain protein residues take part in the reaction while the rest of the enzyme can be described at the MM level. Figure 2.2 shows such a bond crossing the QM-MM border and defines the labels which I will use in the text.

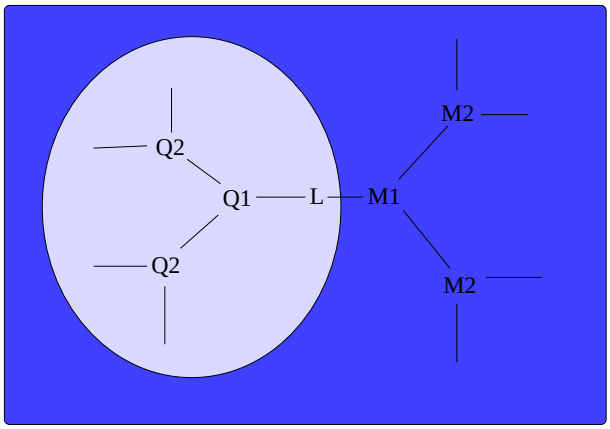


Figure 2.2: A covalent bond across the QM-MM boundary.

2.4.1. Chemical nature of the termination

After the bond is cut one has to deal with three issues:

1. the dangling bond of the QM atom Q^1 must be capped otherwise the system becomes unrealistic.

2. the partial charge of M^1 causes an over-polarization of the QM charge density in the case of electrostatic or polarized embedding.

3. for atoms which are in both subsystems the double-counting of interactions has to be avoided.

If a covalent bond is broken, the most popular approach for the termination is the introduction of a link atom. To saturate the dangling bond of atom Q^1 an additional atom L is introduced, together with basis functions and electrons required to form a covalent bond. The simplest and most popular choice is a hydrogen atom [4]. The problem is that the hydrogen atom has to replace the characteristics of the whole chemical group it replaces. In case of boundary schemes, a special boundary atom replaces the M^1 atom. It is part of both - inner and outer - systems. In the inner part it replaces the cut bond. It is also possible that it replaces the electronic character of the MM part attached to Q^1. In the MM part it behaves like a normal MM atom. Within empirical and semi-empirical schemes the link atom is parameterized to mimic the modeled group. For example Antes and Thiel [29] reported the semi-empirical Adjusted Connection Atom (ACA) scheme. They replace the link atom by a boundary atom and parameterize it to model a methyl group. Another approach in *ab initio* based QM/MM schemes is to place a pseudopotential at the MM site to simulate the electronic properties of the replaced bond [30]. In localized-orbital schemes (frozen) hybrid orbitals are placed at the boundary to cap the QM region, thus replacing the cut bond. While the frozen hybrid orbital approaches promise to mimic the electronic properties of the extended system, problems with the calculation of the $E_{QM-MM}(I,O)$ term in equation (2.3.2) still remain. Philipp and Friesner [31] reported that it is probably necessary to combine the definition of the terminating orbitals with reparameterization of the link atom region to obtain accurate conformational energies.

2.4.2. Coordinates and forces for the termination sites

In this section we will take a closer look at the coordinates and forces of the link atoms. Setting up a structure the link atoms will be placed along the bonds that are terminated. Starting the simulation the geometry of the system is optimized or a molecular dynamics run is performed. Therefore the initial coordinates change and there are different ways of updating the coordinates of the link atoms and handling the forces acting on them.

One approach is the so called optimized link atoms scheme. The coordinates of the link atoms are added to the atom list. They are initially positioned at a certain distance along the Q^1-M^1 bond vector. This adds three structural degrees of freedom for each link atom. During a geometry optimization or a molecular dynamics simulation the coordinates are free to vary [7]. Additional force field terms can be included to favor a position of the link atoms along the M^1-Q^1 bond. For this reason an angle term for Q^1-L-M^1 is introduced with an equilibrium angle of zero degrees [32].

The other approach uses constrained link atom procedures. The positions of the link atoms can be written as functions of the real atom coordinates. It is possible to eliminate them from the set of coordinates used in the optimization or dynamics. This is desirable especially in the case of molecular dynamics simulations. Independent link atom motions would modify the result for example of the evaluation of vibrational frequencies. Maseras and Morokuma [33] first introduced constraints to keep the MM frontier atom M^1 along the Q^1-L bond

and to fix the distances between Q^1 and L and between Q^1 and M^1 using a formulation in internal coordinates. Because this eliminates four, rather than three, degrees of freedom per link atom this formulation was modified by defining the position of the link atoms as a function of the positions of Q^1 and M^1 in Cartesian coordinates:

$$\mathbf{R}_L(\mathbf{R}_{Q^1}, \mathbf{R}_{M^1}) = \mathbf{R}_{Q^1} + g(\mathbf{R}_{M^1} - \mathbf{R}_{Q^1}) \tag{2.4.1}$$

The link atom is placed along the bond between Q^1 and M^1 and the bond length of the link atom to the QM atoms is defined by the distance of the original bond length scaled by a factor g. Most current link atom schemes are based on this definition, while they use different formulations of g. Three degrees of freedom are removed with this formulation. Woo et. al. [34] eliminate the coordinates of the MM atom M^1 rather than the link atom L. The main difference to the formulation in Equation (2.4.1) is that the length of the link bond is determined on the QM level, which allows the link bond to break in principle.

$$\mathbf{R}_{M^1}(\mathbf{R}_{Q^1}, \mathbf{R}_L) = \mathbf{R}_{Q^1} + g'(\mathbf{R}_L - \mathbf{R}_{Q^1}) \tag{2.4.2}$$

Choosing g to be a constant the length of the link bonds varies with the length of the original bond Q^1-M^1. A good choice for g is the ratio of the equilibrium bond lengths of Q^1-L and Q^1-M:

$$g = \frac{d_0(Q^1 - L)}{d_0(Q^1 - M^1)} \tag{2.4.3}$$

The force field parameters can provide the values for d_0.

Expressing the position of a link atom as function of other coordinates makes it transparent to geometry optimizations or molecular dynamics algorithms. However the link atom is part of the QM system and it is treated in the same way as all other QM atoms. The resulting forces, which act on the link atom, are mapped onto the atoms defining the coordinates of the link atom. Propagating the atoms in the next time step, the new position of the link atom is determined by the position of the surrounding atoms rather than using the calculated forces which act on the link atom itself. The QM/MM total energy of the system depends on the coordinates of all atoms $\{\mathbf{R}_i\}$, $i \in S$ of the entire system S and the link atom coordinates $\mathbf{R}_l$, $l \in L$:

$$\tilde{E}_{QM/MM}\left(\{\mathbf{R}_i\}\right) = E_{QM/MM}\left[\{\mathbf{R}_i\}, \mathbf{R}_l\left(\{\mathbf{R}_i\}\right)\right] \tag{2.4.4}$$

Building the derivative of $\tilde{E}$ with respect to the atomic position $\mathbf{R}_k$, $k \in S$, one obtains[3]:

$$\frac{\partial \tilde{E}}{\partial \mathbf{R}_k} = \frac{\partial E}{\partial \mathbf{R}_k} + \frac{\partial E}{\partial \mathbf{R}_l}\frac{\partial \mathbf{R}_l}{\partial \mathbf{R}_k} \tag{2.4.5}$$

If atom k is not involved in the definition of the link atom, the second term in

[3]The QM/MM subscript is dropped to make the equation more 'readable'.

Equation (2.4.5) becomes zero. The derivative $\frac{\partial \mathbf{R}_l}{\partial \mathbf{R}_k}$ is a 3×3 Jacobian matrix and describes the coupling of the link atom and real atom motions as a function of the constraint term. Using the definition of the link atom as it is given in Equation (2.4.1) with the constant g as it is defined in Equation (2.4.3) the Jacobians $\mathbf{J}_{Q^1}$ and $\mathbf{J}_{M^1}$ take a diagonal form:

$$J^{\alpha\beta}_{Q^1} = \frac{\partial R^\alpha_L}{\partial R^\beta_{Q^1}} = (1-g)\delta_{\alpha\beta} \tag{2.4.6}$$

$$J^{\alpha\beta}_{M^1} = \frac{\partial R^\alpha_L}{\partial R^\beta_{M^1}} = g\delta_{\alpha\beta} \tag{2.4.7}$$

$\delta_{\alpha\beta}$ is the Kronecker symbol, α and β are Cartesian components.

If the link atom is placed at a fixed position along the Q^1-M^1 bond, there will be a correction term for both participating atoms. The ChemShell implementation [35] uses this approach as well as the QM-Pot scheme by Sauer et. al. [15].

Finally, one has to consider some adjustment to the force field itself. Following the arguments above, the bending potential of Q^2-Q^1-L effectively provides the force constant for the Q^2-Q^1-M^1 angle bend. One approach is to replace the MM term with the link atom bending potential. But greater accuracy could be achieved by modifying the MM parameters, so they can reflect the presence of the link atom. The same problem has to be considered for torsion angles. In all these methods the Q^1-L distance is kept fix. It has also been suggested to vary this distance to tune the electrostatic characteristics of the termination [35].

In section 2.5 the method will be described in which way the QM/MM scheme is implemented in the CP-PAW program package.

2.5. The QM/MM implementation in the CP-PAW program package

In this section the implementation of the QM/MM in the CP-PAW program package is described in detail. As mentioned before most AMBER implementations use additive schemes. Although AMBER is the force field of choice simulating biological systems with CP-PAW, we use a subtractive scheme in our QM/MM implementation. The advantage is that when we calculate a system, the MM part contains the whole system. In additive schemes the whole system is composed by the inner and the outer system. This leads to a reparameterization of the atoms near the boundary of the subsystems. The only atoms which have to be parameterized in our approach are the dummy atoms. In the MM part the original parameters are valid. For the shadow system which is calculated on the MM level, the dummy atoms disturb the optimal parameterization. But the results of the shadow are subtracted from the entire system and are replaced by the results from the QM calculation. These results are on a higher level of sophistication and thus even better than the results from MM calculations.

Later in Section 4.3.2 the Car-Parrinello method is presented which is the ba-

sic method for molecular dynamics in the CP-PAW program package. Therefore, the QM/MM coupling is formulated by defining a Lagrangian for the coupled system. From the Lagrangian the equations of motion are derived as Euler-Lagrange equations. Energy conservation is ensured, for all approximations enter the Lagrangian and forces, potentials etc, are derived analytically from it.

In the following I will denote the QM system as Q, the entire MM system as A (all) and the inner system on the MM level as S (shadow) with respect to the variables in the program code. The inner system is labeled by C the outer part by E. Thus, the Lagrangian of the inner system can be written as:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}^Q + \mathcal{L}^A - \mathcal{L}^S \\ &= \underbrace{\left[\sum_{i\in C}\frac{1}{2}M_i^Q\left(\dot{\mathbf{R}}_i^Q\right)^2 + \sum_n f_n\langle\dot{\Psi}_n|\tilde{m}_\Psi|\dot{\Psi}_n\rangle - E^Q\left[\mathbf{R}_i^Q, |\Psi_n\rangle, f_n\right]\right]}_{\mathcal{L}^Q} \\ &\quad + \underbrace{\left[\frac{1}{2}\sum_{i\in C+E} M_i^A\left(\dot{\mathbf{R}}_i^A\right)^2 - E^A\left(\mathbf{R}_i^A\right)\right]}_{\mathcal{L}^A} - \underbrace{\left[\sum_{i\in C}\frac{1}{2}M_i^S\left(\dot{\mathbf{R}}_i^S\right)^2 - E_C^M\left(\mathbf{R}_i^S\right)\right]}_{\mathcal{L}^S}\end{aligned} \tag{2.5.1}$$

The Lagrangian depends on the atomic positions $\mathbf{R}_i^Q$ of the central cluster in the QM calculation, the atomic positions $\mathbf{R}_i^S$ of the central cluster in the MM calculation, the atomic positions $\mathbf{R}_i^A$ of the entire system in the MM calculation and their velocities. Even though the atoms are identical in all three subsystems, the masses are treated separately in each subsystem. That arises from the need of reduced masses in the QM system in order to account for the effective masses of the wave functions.

In the next step constraints must be introduced. Although the three Lagrangians are independent and depend on independent sets of coordinates the atoms in the central cluster must have the same coordinates in all three subsystems. Thus, the condition $\mathbf{R}_i^Q = \mathbf{R}_i^S = \mathbf{R}_i^A$ must be fulfilled. The constraint term

$$\Delta\mathcal{L} = \sum_{i\in C}\boldsymbol{\lambda}_i\left(\mathbf{R}_i^Q - \mathbf{R}_i^A\right) + \sum_{i\in C}\boldsymbol{\rho}_i\left(\mathbf{R}_i^Q - \mathbf{R}_i^S\right) \tag{2.5.2}$$

with the Lagrange parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ is added to the Lagrangian. This leads to the Euler-Lagrange equations:

$$M_i^Q\ddot{\mathbf{R}}_i^Q = -\nabla_{\mathbf{R}_i^Q}E^Q + \boldsymbol{\lambda}_i + \boldsymbol{\rho}_i \tag{2.5.3}$$

$$M_i^A\ddot{\mathbf{R}}_i^A = -\nabla_{\mathbf{R}_i^A}E^A + \boldsymbol{\lambda}_i \tag{2.5.4}$$

$$-M_i^S\ddot{\mathbf{R}}_i^S = +\nabla_{\mathbf{R}_i^S}E^S - \boldsymbol{\rho}_i \tag{2.5.5}$$

The Lagrange parameters $\boldsymbol{\lambda}_i$ and $\boldsymbol{\rho}_i$ have to be chosen in a way that the constraints are exactly fulfilled in the following time step. This requires:

$$\mathbf{R}_i^Q = \mathbf{R}_i^A = \mathbf{R}_i^S \tag{2.5.6}$$

We can assume that the masses of the atoms in the shadow and in the environment subsystem are pairwise identical, i.e. $M_i^A = M_i^S$. Following this assumption the equations of motion (2.5.3), (2.5.4) and (2.5.5) can be added which leads to:

$$M_i^Q \ddot{\mathbf{R}}_i^Q = -\nabla_{\mathbf{R}_i^Q} E^Q - \nabla_{\mathbf{R}_i^A} E^A + \nabla_{\mathbf{R}_i^S} E^S \tag{2.5.7}$$

So far link bonds have been ignored. Link bonds have been introduced in Section 2.4. In the following the atoms participating in a link bond are denoted by indices I and O. The inner atom (I) is the same atom in each subsystem. The outer atom (O) is a dummy hydrogen atom in the QM (Q) and the shadow (S) subsystem and the original atom in the MM (A) subsystem. The constraint term in the Lagrangian is replaced by

$$\begin{aligned} \Delta\mathcal{L} &= \boldsymbol{\lambda}_O \left(\mathbf{R}_O^Q - \left(\mathbf{R}_I^A + \alpha \left[\mathbf{R}_O^A - \mathbf{R}_I^A \right] \right) \right) + \boldsymbol{\lambda}_I \left(\mathbf{R}_I^Q - \mathbf{R}_I^A \right) \\ &\quad + \boldsymbol{\rho}_I \left(\mathbf{R}_O^Q - \mathbf{R}_O^S \right) + \boldsymbol{\rho}_I \left(\mathbf{R}_I^Q - \mathbf{R}_I^S \right) \end{aligned} \tag{2.5.8}$$

The factor α determines the position of the dummy atom along the link bond. It is defined by:

$$\alpha = \frac{R_S + R_J}{R_M + R_J} \tag{2.5.9}$$

R_S is the covalent radius of the dummy atom in the shadow system, R_J is the covalent radius of the inner atom in the MM system and R_M is the covalent radius of the outer atom in the MM system. From the Lagrangian for the atoms participating in link bonds a new set of Euler-Lagrange equations can be derived. The Lagrange parameters are defined by the constraint conditions:

$$\mathbf{R}_I^Q = \mathbf{R}_I^A = \mathbf{R}_I^S \qquad \text{and} \qquad \mathbf{R}_O^Q = \mathbf{R}_I^A + \alpha \left(\mathbf{R}_O^A - \mathbf{R}_I^A \right) = \mathbf{R}_O^S \tag{2.5.10}$$

The Lagrange parameters are determined in the following way: First, the link bonds are propagated without constraints. In a second step the constraints are applied to the link bonds. This is done in the subroutine QMMM$PROPAGATE, while all non-link bond participating atoms are propagated in QMMM$INTERFACE. Two new symbols have to be introduced at this point: $\mathcal{R}$ stands for the positions propagated without constraint forces and $\chi = \frac{\Delta^2}{(1+a)M}$ where $a = \frac{\gamma\Delta}{2}$, Δ is the time step, γ is a friction constant. The equations of motion are propagated using the Verlet algorithm which is described in Equation (4.3.6) in Section 4.3.1. The constraint conditions are:

$$\begin{aligned} \mathcal{R}_I^Q + (\boldsymbol{\lambda}_I + \boldsymbol{\rho}_I)\chi_I^Q &= \mathcal{R}_I^A - ((1-a)\boldsymbol{\lambda}_O + \boldsymbol{\lambda}_I)\chi_I^A \\ \mathcal{R}_I^Q + (\boldsymbol{\lambda}_I + \boldsymbol{\rho}_I)\chi_I^Q &= \mathcal{R}_I^S + \boldsymbol{\rho}_I \chi_I^S \\ \mathcal{R}_O^Q + (\boldsymbol{\lambda}_O + \boldsymbol{\rho}_O)\chi_O^Q &= \mathcal{R}_O^S + \boldsymbol{\rho}_O \chi_O^S \\ \mathcal{R}_O^Q + (\boldsymbol{\lambda}_O + \boldsymbol{\rho}_O)\chi_O^Q &= (1-\alpha)\left[\mathcal{R}_I^A - ((1-a)\boldsymbol{\lambda}_O + \boldsymbol{\lambda}_I)\chi_I^A \right] \\ &\quad + \alpha \left[\mathcal{R}_O^A - a\boldsymbol{\lambda}_O \chi_O^A \right] \end{aligned}$$

These equations define a system of linear equations for the Lagrange parameters. It can be rewritten as:

$$\begin{aligned}
(1-\alpha)\chi_I^A\boldsymbol{\lambda}_O + \left(\chi_I^Q + \chi_I^A\right)\boldsymbol{\lambda}_I + \chi_I^Q\boldsymbol{\rho}_I &= \mathcal{R}_I^A - \mathcal{R}_I^Q \\
\chi_I^Q\boldsymbol{\lambda}_I + \left(\chi_I^Q - \chi_I^S\right)\boldsymbol{\rho}_I &= \mathcal{R}_I^S - \mathcal{R}_I^Q \\
\chi_O^Q\boldsymbol{\lambda}_O + \left(\chi_O^Q - \chi_O^S\right)\boldsymbol{\rho}_O &= \mathcal{R}_O^S - \mathcal{R}_O^Q \\
\left[\chi_O^Q + (1-\alpha)^2\chi_I^A + \alpha^2\chi_O^A\right]\boldsymbol{\lambda}_O + (1-\alpha)\chi_I^A\boldsymbol{\lambda}_I + \chi_O^Q\boldsymbol{\rho}_O &= (1-\alpha)\mathcal{R}_I^A + \alpha\mathcal{R}_O^A - \mathcal{R}_O^Q
\end{aligned}$$

The following code fragment shows how the system of equations is solved in the program:

```
CHIQO=1.D0/(MASS(IATQO)*(1.D0+ANNERVEC(IATQO)))
CHIQI=1.D0/(MASS(IATQI)*(1.D0+ANNERVEC(IATQI)))
CHIAO=1.D0/(MA(IATAO)*(1.D0+ANNER))
CHIAI=1.D0/(MA(IATAI)*(1.D0+ANNER))
CHISO=1.D0/(MS(IATSO)*(1.D0+ANNER))
CHISI=1.D0/(MS(IATSI)*(1.D0+ANNER))
MAT(:,:)=0.D0
MAT(1,1)=(1.D0-ALPHA)*CHIAI
MAT(1,2)=CHIQI+CHIAI
MAT(1,4)=CHIQI
MAT(2,2)=CHIQI
MAT(2,4)=CHIQI-CHISI
MAT(3,1)=CHIQO
MAT(3,3)=CHIQO-CHISO
MAT(4,1)=CHIQO+(1.D0-ALPHA)**2*CHIAI+ALPHA**2*CHIAO
MAT(4,2)=(1.D0-ALPHA)*CHIAI
MAT(4,3)=CHIQO
CALL LIB__INVERTR8(4,MAT,MATINV)
DO I=1,3
  VEC(1)=RAIP(I)-RP(I,IATQI)
  VEC(2)=RSIP(I)-RP(I,IATQI)
  VEC(3)=RSOP(I)-RP(I,IATQO)
  VEC(4)=(1.D0-ALPHA)*RAIP(I)+ALPHA*RAOP(I)-RP(I,IATQO)
  LAGR(:)=MATMUL(MATINV,VEC)
  RP(I,IATQO)=RP(I,IATQO)+CHIQO*(LAGR(1)+LAGR(3))
  RP(I,IATQI)=RP(I,IATQI)+CHIQI*(LAGR(2)+LAGR(4))
  RAOP(I)=RAOP(I)+CHIAO*(-ALPHA*LAGR(1))
  RAIP(I)=RAIP(I)+CHIAI*(-(1.D0-ALPHA)*LAGR(1)-LAGR(2))
  RSOP(I)=RSOP(I)+CHISO*LAGR(3)
  RSIP(I)=RSIP(I)+CHISI*LAGR(4)
ENDDO
```

Finally, the electrostatic coupling between the subsystems has to be discussed. In the environment the point charges are used, which are provided by the topology file. In the QM subsystem the quantum mechanical charge density is known. It has to be transformed into atom-centered point charges in order to use the charges in the force field calculations of all subsystems. This transformation is done by fitting the charge density to a superposition of Gaussians which are centered at the atom sites.

Using the point charges, the electrostatic energy of the combined system is given by:

$$E_{C+E} = \sum_{i,j\in C+E, i\neq j} \frac{Q_i Q_j}{8\pi\varepsilon_o|\mathbf{R}_i - \mathbf{R}_j|} \tag{2.5.11}$$

The variation of the energy with respect to the point charges yields the point potentials:

$$V_{C+E,i} = \frac{\partial E_{C+E}}{\partial Q_i} = \sum_{j \in C+E, j \neq i} \frac{Q_j}{4\pi\varepsilon_0 |\mathbf{R}_i - \mathbf{R}_j|} \tag{2.5.12}$$

This potential is included in the potential for the electrons in the central cluster. Thus, the environmental point charges can polarize the quantum mechanical charge distribution.

The total charge density is given as:

$$\rho^A(\mathbf{r}) = \sum_{i \in C+E} Q_i^{A,0} \delta(\mathbf{r} - \mathbf{R}_i^A) + \sum_{i \in C} Q_i^Q \delta(\mathbf{r} - \mathbf{R}_i^Q) - \sum_{i \in C} Q_i^{S,0} \delta(\mathbf{r} - \mathbf{R}_i^S) \tag{2.5.13}$$

The superscript '0' denotes charges which are given from the forcefield parameters. The charges with the superscript 'Q' are constructed from the charge density of the QM subsystem.

Again the atoms involved in link bonds need special treatment. Therefore, the charge correction of the dummy atom is distributed to the real atoms of the link bond. Starting from the conditions

$$\Delta Q_O^A + \Delta Q_I^A = \Delta Q_D \quad \text{and} \quad \Delta Q_O^A \mathbf{R}_O^A + \Delta Q_I^A \mathbf{R}_I^A = \Delta Q_D \mathbf{R}_O^Q \tag{2.5.14}$$

the charge correction is given by:

$$Q_O^A = Q_O^{A,0} + \Delta Q_O^{A,0} = Q_O^{A,0} + \alpha \left(Q_O^Q - Q_O^S \right) \tag{2.5.15}$$

$$Q_I^A = Q_I^{A,0} + \Delta Q_I^{A,0} = Q_I^{A,0} + (1-\alpha) \left(Q_O^Q - Q_O^S \right) \tag{2.5.16}$$

ΔQ_D is the charge correction of the dummy atom and is defined by $\Delta Q_D = Q_O^Q - Q_O^S$. The splitting of the dummy atom's charge creates an unphysical self energy. This can be corrected by subtracting the change of the electrostatic energy of the isolated link bond. The correction term is given by:

$$\Delta E = -\frac{e^2 \Delta Q_O^A \Delta Q_I^A}{4\pi\varepsilon_0 |\mathbf{R}_O^A - \mathbf{R}_I^A|} \tag{2.5.17}$$

For the correction depends on the positions of the atoms, it also leads to a correction of the forces. However, in the present version of CP-PAW the correction term has not been included yet.

The charge distribution of each subsystem (QM and MM) affects the electrostatic energy of the other subsystem. The potential is corrected by:

$$v(\mathbf{r}) = \frac{\partial E}{\partial n(\mathbf{r})} = \frac{\partial E^Q}{\partial n(\mathbf{r})} + \sum_{i \in C} \left(\frac{\partial E^A}{\partial Q_i^Q} - \frac{\partial E^S}{\partial Q_i^Q} \right) \frac{\partial Q_i^Q}{\partial n(\mathbf{r})} \tag{2.5.18}$$

The electrostatic potentials acting on the point charges can be defined by:

$$V_i^A := \frac{\partial E^A}{\partial Q_i^Q}; \qquad V_i^S := \frac{\partial E^S}{\partial Q_i^Q}; \qquad V_i^Q := \frac{\partial E^A}{\partial Q_i^Q} - \frac{\partial E^S}{\partial Q_i^Q} \tag{2.5.19}$$

It can be easily seen that for atoms of the central cluster, which do not participate in link bonds, the potentials are given as:

$$V_i^Q = V_i^A - V_i^S \tag{2.5.20}$$

In the case of link bonds the potentials of the inner and outer atoms are corrected by:

$$V_O^Q = \alpha V_O^A + (1-\alpha) V_I^A - V_O^S \tag{2.5.21}$$

$$V_I^Q = V_I^A - V_I^S \tag{2.5.22}$$

3. Force field simulations

3.1. Force Fields

3.1.1. Review of Current Force Fields

Since the first force fields were developed, two main branches have been established.

The primary goal of the first branch is to make accurate predictions of molecular structures and properties. Allinger et. al. developed the early members of this family, the so called Molecular Mechanics Force Fields, MM1 [36] and MM2 [37]. These first force fields have been superseded by the improved models MM3 and MM4. The new models have been parameterized against experimental data like heats of formation and vibrational frequencies. A complicated functional form is needed to reproduce the experimental data accurately. The MM3 force field [38, 39, 40] employs a total of nine interactions like stretch-bend, stretch-torsion or bend-bend interactions (depicted in Figure 3.1) additionally to the common interactions used by all force fields (which will be discussed in section 3.1.2). The MM4 force field [41, 42, 43, 44] goes one step further and adds

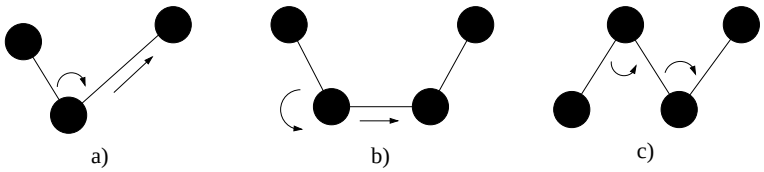


Figure 3.1: The stretch-bend interaction (a), the stretch-torsion interaction (b) and the bend-bend interaction (c) in the MM3 force field.

additionally six interactions to the energy functional. These are stretch-stretch, torsion-bend, bend-torsion-bend, torsion-torsion, torsion-improper torsion and improper torsion-torsion-improper torsion interactions. Also the van-der-Waals potential is modeled using the Hill potential as it is defined in equation (3.1.13).

The second branch aims at modeling large molecules such as proteins and polymers. The most common methods in this category are the AMBER force field [45, 21], CHARMM [20] and the OPLS force fields [46]. These force fields have a simpler functional form, only using the basic interactions. Bond stretching potentials and bending potentials are restricted to a harmonic approximation, the van-der-Waals interaction is modeled by a Lennard-Jones potential. These force fields often give good results for condensed phase properties, despite their simplicity.

The Merck Molecular Force Field (MMFF) [47, 48, 49, 50, 51] is an example of a force field which tries to take advantage of both branches. It uses one cross

term to describe bend-stretch interaction and uses quartic and cubic expressions for bond stretching and bending. Thus it lies in its complexity between AMBER and MM3/MM4 force fields. Unlike most other force fields it has not been parameterized from empirical data but from *ab initio* calculations. Halgren showed that this approach performs even better than the MM3 force field in case of structures of small organic molecules [52]. However, it performs poorly in condensed phase calculations, because condensed phase are neglected in the parameterization.

3.1.2. The functional form of a force field

In this section the construction of the functional form of a force field is discussed. Many force fields are based on five basic interactions which take part in the total energy. Special interactions, for example the ones employed by the MM3 and MM4 force fields, will not be discussed in this thesis. The AMBER force field, which is used during the work on this thesis, only uses the basic interactions.

The total energy of a system is a function of the atom coordinates. The interactions between the atoms can be divided into two groups: the bonded interactions and the non-bonded interactions. Bond stretching, angle interactions and torsion interactions belong to the first group. Non-bonded interactions are the Coulomb and the van-der-Waals interactions between the atoms.

The total energy is given by:

$$E(r) = E_{bond} + E_{angle} + E_{torsion} + E_{coulomb} + E_{vdW} \tag{3.1.1}$$

In the following each term will be described in detail.

Bond stretching

Consider two atoms which form a bond. The energy of this system depends on the distance r between the two atoms and has a minimum. This minimum is called bond energy and the distance is called the bond length; it is the equilibrium distance r_0 of the two atoms. To construct a functional form of the bond energy, one has to consider how the energy changes with the bond length.

If the bond is compressed the electron clouds of the two atoms will overlap which results in a strong increase in energy. Stretching the bond beyond its equilibrium length the bond energy increases until the atoms dissociate. This is plotted in Figure 3.2 by the black line. Around the minimum r_0 the energy can be written as a Taylor expansion in $r - r_0$:

$$\begin{aligned} E(r) &= E(r_0) + \left.\frac{\mathrm{d}E}{\mathrm{d}r}\right|_{r=r_0} (r - r_0) + \frac{1}{2} \left.\frac{\mathrm{d}^2E}{\mathrm{d}r^2}\right|_{r=r_0} (r - r_0)^2 \\ &\quad + \mathcal{O}\left((r - r_0)^3\right) \end{aligned} \tag{3.1.2}$$

Terminating the Taylor expansion after the quadratic term is the common method in many force fields. This approximation is called harmonic approximation. In the equilibrium distance[1] r_0 the force and hence the first derivative

[1] Talking about the *equilibrium* distance can be misleading. It is the *reference* bond length

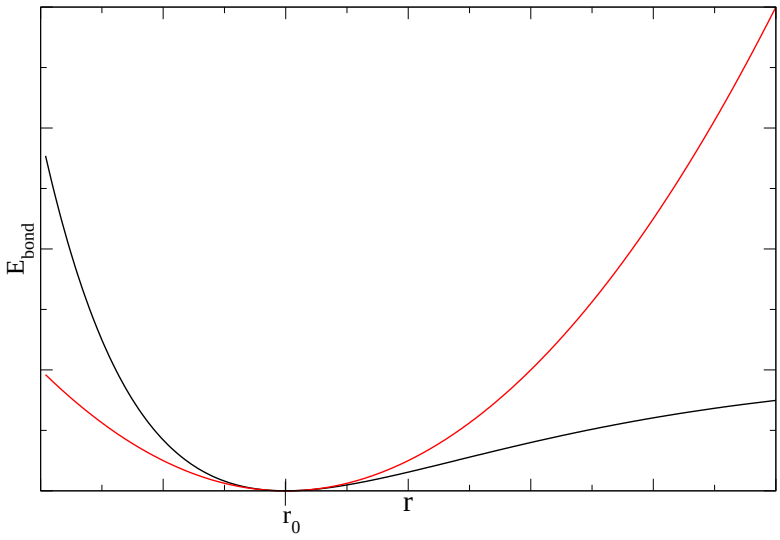


Figure 3.2: Potential energy of a bond between two atoms. The black line shows the energy of the stretched bond. It has a minimum at r_0. For $r \longrightarrow 0$ the energy increases to infinity, for $r \longrightarrow \infty$ the energy approximates the dissociation energy. The red line shows the harmonic approximation used by many force fields. The energy has been determined using Equation (3.1.3).

of the energy is zero. Setting $E(r_0) = 0$ the bond potential energy is:

$$E_{bond}(r) = \frac{1}{2} k_r (r - r_0)^2 \tag{3.1.3}$$

k_r is the harmonic force constant which is defined by

$$k_r = \left. \frac{\mathrm{d}^2 E}{\mathrm{d}r^2} \right|_{r=r_0} \tag{3.1.4}$$

The red line in Figure 3.2 shows the energy with respect to the bond length in the harmonic approximation. This approximation can be justified because the forces between bonded atoms are higher than the forces by other interactions. Large deviations from the equilibrium bond length lead to large deviations from the true energy. If bond lengths far from r_0 are needed in a simulation higher order terms of the Taylor expansion have to be included into the energy expression. Finally, an error remains for atoms which are separated by a large distance. $E(r)$ goes to infinity when $r \longrightarrow \infty$, which is an unphysical behavior. Therefore, some force fields employ a Morse potential to describe the bond energy:

$$E_{bond}(r) = D \left(1 - \exp(-\alpha(r - r_0))\right)^2 \tag{3.1.5}$$

when all other force field terms are set to zero. The true *equilibrium* bond length is the bond length when the molecule is in its minimum energy configuration.

where D is the dissociation energy and α is given by $\alpha = \sqrt{\frac{k}{2D}}$. The consequence is that harmonic expressions cannot describe the breaking of bonds, while the Morse potential is able to do that.

Bond bending

It is a common method to use a harmonic potential for bond bend terms like the way it is defined in AMBER:

$$E_{angle}(\theta) = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{3.1.6}$$

k_θ is the angle bending force constant, θ is the actual bond angle, while θ_0 describes the equilibrium bond angle. The accuracy can be increased by including higher order terms in the expression. The energy to disturb an angle from its equilibrium is much lower than disturbing a bond. Therefore, the force constants k_θ are smaller than those for bond stretching.

While AMBER uses the simple harmonic expression for the angle bend term other force fields use a Fourier expansion. Landis showed that harmonic angular potential energy terms have the unwanted properties of overestimating bond angle energies at angles far from the equilibrium position [53]. He uses

$$E_\theta = k_F \left[1 + \cos(n\theta + \psi)\right] \tag{3.1.7}$$

in the SHAPES method, where n is the periodicy of the cosine function and ψ is the phase shift which determines the positions of the minima. k_F is the force constant in the Fourier expression. The relationship between the harmonic force constant k_θ and k_F is given by:

$$k_F = 2k_\theta / n^2 \tag{3.1.8}$$

Figure 3.3 shows the value and the derivative of the angular potential for HC-CT-C around the equilibrium.

Torsion angle interactions

Although early force fields omitted the torsional interactions, they are the important terms to describe the intramolecular interactions. They have a large part in determining the molecular structure, especially the secondary structure of proteins. However, they are different than the described bond and angle interactions. First, the rotational barriers are low compared to the other discussed interactions. This means that the dihedral angles can make large changes. Secondly, the torsional potential E_{tors} is periodic through a 360° rotation. A harmonic expression would not describe this behavior properly. Therefore, a Fourier series is used to model the torsional interaction:

$$E_{torsion}(\theta) = \sum_n \frac{1}{2} V_n \cos(n\theta) \tag{3.1.9}$$

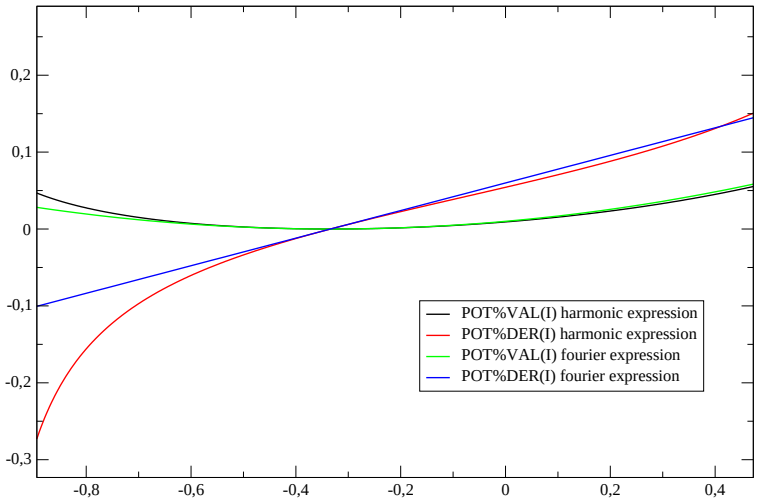


Figure 3.3: Value and derivative of the angular potential in a harmonic and in a Fourier expression. The potential is shown for HC-CT-C (THETA=109., K=50.). Around the equilibrium angle both potentials have the same value and derivative.

V_n are the force constants of the torsional rotation, θ is the current torsional angle. In many force fields the zero of the potential is shifted and a phase factor is introduced. This results in:

$$E_{torsion}(\theta) = \sum_n \frac{1}{2} V_n \left(1 + \cos\left(n\theta + \theta_0\right)\right) \tag{3.1.10}$$

The phase factor θ_0 is chosen in a way that terms with a positive V_n have a minimum at $\theta = 180°$[2]. Figure 3.4 shows the first three terms of the Fourier expansion.

At this stage the main terms of bonded interactions have been discussed. What is left are the non-bonded interactions.

Non-bonded interactions

Non-bonded interactions will occur in two cases. They can act on atoms which are in the same molecule and on atoms which belong to different molecules. Most force fields use two different interactions: the electrostatic interaction and the van-der-Waals interaction.

Consider a water molecule. The water molecule is a dipole with the negative charge located at the oxygen atom. Within a force field model the charge distribution will be modeled by atomic point charges. To ensure charge conservation the total charge of the system has to be zero. In the case of the water molecule the charges have to be $-q_O = q_{H_1} + q_{H_2}$. The energy term for the interaction

[2] This leads to $\theta_0 = 0°$ in case of odd n and $\theta_0 = 180°$ in case of an even n.

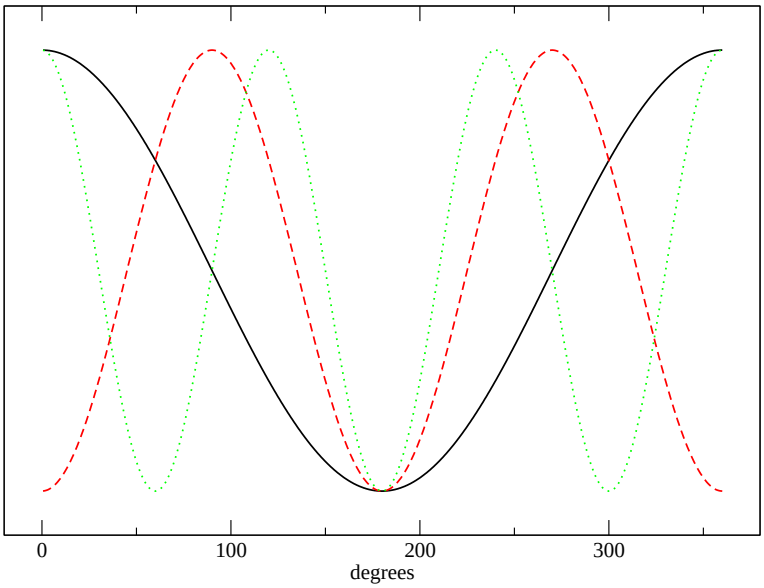


Figure 3.4: The first three terms form equation (3.1.10). The bold and black line shows $n = 1$, the red and dashed line shows $n = 2$ and the green and dotted line shows the $n = 3$ term.

between the point charges is given by the Coulomb potential:

$$E_{coulomb} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{3.1.11}$$

ε_0 is the vacuum permittivity, q_i and q_j are the atomic point charges and r_{ij} is the distance between the atoms i and j.

All non-bonded interactions between atoms which are not electrostatic interactions are brought together in the van-der-Waals interaction. This includes dispersion and repulsion as well as some other interactions [54]. Dispersion arises due to correlations between electrons in different atoms. The energy is lowered and therefore this interaction has an attracting character. The repulsion comes from the interaction between two electron clouds. If they overlap the Pauli repulsion leads to a rise in energy.

To find the functional form of E_{vdW} one has to consider its behavior with respect to the distance. Small distances between two atoms lead to strong repulsive forces. The energy has a large positive value, approximating infinity as the distance goes to zero. For larger distances the dispersion interaction outbalances the repulsion. The energy reaches small negative values which approximate zero when the distance becomes infinity. One of the most common forms for E_{vdW} is the Lennard-Jones (LJ) potential, which has exactly the above

described behavior. It is defined by

$$E_{vdW}^{LJ} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{3.1.12}$$

A_{ij} and B_{ij} are suitable parameters which ensure that the potential has its minimum at the equilibrium distance of the atoms i and j and that the dissociation energy has the right value. The choice of the r^{-12} behavior of the repulsive term is not based on physical arguments but is convenient due to computational advantages.

Some force fields like MM3 and MM4 choose an exponential term for the repulsive part. The repulsive potential is dominated by the overlap between electron clouds on different atoms. It follows that the potential should have an exponential dependence with distance, as the electron density falls off exponentially with distance from the nuclei. The Hill potential [55] is one example for an exponential repulsive term:

$$E_{vdW}^{Hill} = A_{ij} \exp\left[-B_{ij} r_{ij}\right] - \frac{C_{ij}}{r_{ij}^{6}} \tag{3.1.13}$$

A_{ij}, B_{ij} and C_{ij} are constants. The Hill potential retains the r^{-6} dependence of the Lennard-Jones potential and is also called a 6-exponential representation. As the distance approximates zero the Hill potential goes through a maximum and falls to minus infinity. This leads to instabilities if the distance of two atoms becomes too small. Therefore, a cutoff radius has to be introduced.

Non-bonded exclusions

For all non-bonded interactions it is important to define which atoms interact with each other and which atoms do not. The Coulomb and the van-der-Waals interaction lead to instabilities if the distance between the interacting atoms becomes too small. Setting up the force field an exclusion list has to be created. Atoms which are connected by a bond (1-2 bonded) and atoms which are bonded by a common atom (1-3 bonded) are excluded from the non-bonded interactions. The treatment of 1-4 bonded atoms differs from force field to force field. While AMBER excludes these atoms from the non-bonded interactions, UFF takes them into account. The non-bonded interactions of 1-4 bonded atoms are often scaled, in order to prevent the non-bonded interaction to become larger than the torsional interaction.

3.2. The AMBER force field

As it is used today, the AMBER force field[3] is a second generation force field developed by Cornell et. al. [21]. It is based on the original AMBER force field

[3] Through the whole thesis the Cornell et. al. force field is meant when talking about the AMBER force field. AMBER is the name of the commercial program package which originally used the Weiner et. al. force field.

developed by Weiner et. al. in 1984 [45]. The force field was developed for proteins and nucleic acids and it proved to be quite effective [56, 57]. However, the biomolecules could only be simulated in vacuum. The computer power in these days did not allow any explicit solvent. With growing computer power explicit solvent representations were included into force fields for biomolecules. The first method came from Jorgensen and co-workers. Their Optimized Potentials for Liquid Simulations model (OPLS) [58] combine non-bonded parameters, fitted for the description of the solvent, with the bonded parameters from the AMBER force field. This OPLS/AMBER approach for peptides and proteins has been successfully used in many systems [59, 60].

The AMBER force field of the second generation has been influenced by the OPLS method. The OPLS charges have been derived empirically, based on the properties of liquids. The OPLS force field parameterizes the atomic charges by an electrostatic potential (ESP) fit derived with (presumably) Hartree-Fock calculations using the 6-31G* basis set [61]. This basis set is known for overestimating molecular polarity. This was a logical choice because the common water models (e.g. TIP3P, TIP4P [62]) have dipole moments that are about 20% higher than the gas-phase values for water [63]. Nevertheless, the 6-31G* based ESP-fit charge model is not an ideal choice. First, there is a variation of charges if they are derived using different conformations of a molecule. Problems arise when generating other force field parameters with these charges. Secondly, there are charges of atoms which are shielded or buried by other atoms (e.g. sp3 carbon atoms in butane). These charges are statistically under-determined and often assume unexpectedly large values for nonpolar atoms. These deficiencies motivated Bayly et. al. to develop the restrained ESP-fit model (RESP) [64]. The RESP model involves a least-square fit of the electrostatic charges as known from the OPLS approach. But in addition it uses hyperbolic restraints on charges on non-hydrogen atoms. With these restraints the charges of buried atoms can be reduced without impacting the whole fit. This charge model has proven to perform well at reproducing interaction energies and free energy of solvation. It also results in good conformational energies for many small molecules [65]. In the Weiner et. al. force field the ESP fit of the atomic charges have been derived by quantum mechanical calculations.

Reproducing the properties of liquids requires not only fitted charges but also the van-der-Waals (vdW) parameters have to be considered. The vdW parameters in the Weiner et. al. force field were a modification of a parameter set proposed by Hagler, Euler and Lifson [66]. They fitted the vdW parameters to lattice energies and crystal structures of amides. In the Cornell et. al. force field the vdW parameters were improved to describe the ϕ and ψ dihedral parameters of peptide backbones properly. The parameterization is based on quantum mechanical data of the conformational energies of the glycyl and alanyl dipeptides[4][67, 68]. In the Weiner et. al. force field the torsional parameters of the ϕ and ψ angles were set to 0.0 kcal/mol because no theoretical data were

[4] These dipeptides are the basis for the secondary structure of proteins. Alanyl dipeptides are quite rare but they can be found in γ-turns. Glycyl dipeptide are the conformations found in β-sheets. The vdW parameters are fitted in the way that the error in conformational energy is small.

available at this time to fit the parameters.

3.2.1. Details of the model

The functional form of the AMBER model is quite simple. It only depends on five basic interactions which have been described in section 3.1.2. It uses a simple harmonic expression for the bond and the angle terms. The torsional energy is built by a simple set of parameters. In many cases only the two central atoms specify the torsion. The vdW energy is built by a 6-12-potential, the electrostatic interaction is modeled by a Coulomb interaction of atomic point charges. The non-bonded interactions are calculated for atoms which belong to different molecules or which are separated by at least three bonds. For those which are separated by exactly three bonds (1-4 bonds) the non-bonded interactions are reduced by a scale factor.

Thus the total energy is given by:

$$\begin{aligned} E_{total} &= \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \\ &+ \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\phi - \gamma)\right] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] \end{aligned} \tag{3.2.1}$$

Details of the implementation of the AMBER force field can be found in Section 5.1.5. In the following, the model is described in a more general way.

Cornell and co-workers assume that this simple representation is good enough for modeling most unstrained systems. The main goal of the force field is to accurately model conformational energies and intermolecular interactions of proteins and nucleic acids. The parameters can be easily expanded for other molecules which have similar functional groups.

The atom types provided by the force field are listed in [21]. The list contains only a small variety of elements. This is because the force field concentrated on biomolecules. Only hydrogen, carbon, oxygen, nitrogen, sulfur and phosphorous have been parameterized. However, there is more than one force field type of each element, due to the chemical surrounding of an atom.

The bond and angle parameters were taken from the Weiner et. al. force field as starting values. These values have been adjusted to reproduce experimental normal mode frequencies. Structural and vibrational frequencies data have been used to derive the parameters. For example, in complex fragments such as the nucleic acid bases, the r_{eq} and θ_{eq} values have been taken from X-ray structural data. The plausibility of the parameters has been proven by Halgren [69].

The parameters for the dihedrals were developed by Weiner et. al. Their set was created to reproduce the functionalities found in proteins and DNA. It was calibrated to adjust the energies of small model compounds. This is a quite different approach from what is used by many other force fields use. Other parameter sets are optimized to best reproduce the conformational energies of a large number of molecules. In the AMBER force field a dihedral parameter is optimized for the simplest molecule it occurs. Then it is applied to larger systems. The advantage of this approach is the lack of dependence of the resulting

parameters on the set of molecules for testing them.

A large part of the van-der-Waals parameter set has been taken from the OPLS force field. However, some parameters were optimized. The vdW parameters of sp^3 carbon atoms and aliphatic hydrogen were developed using Monte Carlo simulations[5] of CH_4, C_2H_6, C_3H_8 and C_4H_{10} liquids. Additional simulations of liquid benzene were made to fix the parameters for sp^2 carbon atoms. In the original OPLS force field these parameters were not available. The TIP3P water model is used in the Cornell et. al. force field. It is described in detail in section 3.2.2. Therefore the R^* and ϵ are zero for the hydrogen atoms (HW) of a water molecule.

As mentioned above the force field model uses a 6-12 vdW representation and not a 6-exponential representation as shown in equation (3.1.13). The lack of polarization of the force field model and the used 6-12 representation lead to exaggerated short-range repulsions. Therefore the 1-4 vdW interactions have to be scaled down. A scale factor of 1/2.0 as used in the original Weiner et. al. force field has been retained.

The point charges for the electrostatic model have been derived using the 6-31G* basis set and the RESP fitting approach. These models have proven to give good results in calculating the liquid enthalpies and densities [70] and free energies of solvation [65] of methanol and N-methyl acetamide (NMA)[6] in good agreement with the experiment. The charge model employs a scale factor of 1/1.2 for 1-4 electrostatic interactions.

3.2.2. The water model in AMBER - TIP3P

Simulations of protein structure in the gas phase cannot be compared to most experimental data because the solvent has a large effect in organic- and biochemistry. Especially when studying DNA and proteins the solvent is responsible for the structure and thus for the functionality of the molecules. A theoretical model is needed which describes the properties of liquids and solutions. The basis of such a model is an intermolecular potential function which describes the interaction between monomers in the fluid. The Cornell et. al. force field uses the TIP3P water model as it has been developed by Jorgensen et. al. [71, 72, 62]. Jorgensen started his work by searching for transferable intermolecular potential functions (TIPS). The TIPS should be as simple as possible but should give reasonable structural and energetic results for simple liquids (i.e. water, alkanes, alcohols and ethers) and trends for series of dimers. The potentials are defined by the number of interaction sites for a monomer. This number squared gives the number of interatomic distances which have to be evaluated, thus, determining the computational effort.

The interaction energy is calculated between all intermolecular pairs of sites. It consists of a 12-6 Lennard-Jones potential and a Coulomb potential. The

[5] No more detail of these simulations have been presented in the literature.
[6] Methanol and NMA were used as prototypes for polar molecules.

	TIP3P	TIP4P
r_{OH} in Å	0.9572	0.9572
$\angle HOH$ in degree	104.52	104.52
$A \cdot 10^{-3}$ in kcal Å^{12}/mol	582.0	600.0
C in kcal Å^6/mol	595.0	610.0
q_O	-0.834	0.0
q_H	0.417	0.52
q_M	0.0	-1.04
r_{OM} in Å	0.0	0.15

Table 3.1: Geometry data and parameters for potential functions of the water monomer in TIP3P and TIP4P water models. Data taken from [62].

interaction energy ϵ_{mn} between monomers m and n is defined by:

$$\epsilon_{mn} = \sum_{i}^{\text{on } m} \sum_{j}^{\text{on } n} \left(\frac{q_i q_j e^2}{r_{ij}} + \frac{A_i A_j}{r_{ij}^{12}} - \frac{C_i C_j}{r_{ij}^6} \right) \tag{3.2.2}$$

The sums run over all sites of one monomer. Each site has three parameters (q, A and C) which have to be determined so that they yield the right structural and energetic results. Hagler et. al. found that the Lennard-Jones parameters of the hydrogen atom can be set to zero, still yielding accurate results [73]. The remaining parameters are A_O, C_O and q_O (the charge of the hydrogen atoms is defined by the oxygen charge: $q_H = q_O/2$). Equation (3.2.2) is therefore reduced to:

$$\epsilon_{mn} = \sum_{i}^{\text{on } m} \sum_{j}^{\text{on } n} \left(\frac{q_i q_j e^2}{r_{ij}} \right) + \frac{A_O}{r_{OO}^{12}} - \frac{C_O}{r_{OO}^6} \tag{3.2.3}$$

It is important to note that the TIPS models are based on rigid monomers.

The parameters used in the TIP3P model are listed in Table 3.1. The table also contains the parameters for the TIP4P. The TIP4P is a 4-site model which proved to yield better results for water. The idea behind the 4-site model is to move the negative charge away from the oxygen model towards the hydrogens. The charge is placed at a point M which lies on the bisector of the HOH angle.

The accuracy of water models is evaluated by comparing the results to experimental data. Important tests are the geometry of a water dimer and the radial distribution functions (RDF). RDFs count the number of atoms in a given distance. In the case of water three RDFs are of interest: the number of oxygen atoms in a distance to one oxygen atom $g(OO)$, the number of hydrogen atoms in a distance to one oxygen atom $g(OH)$ and the number of hydrogen atoms in a distance to one hydrogen atom $g(HH)$. These functions can be measured with x-ray diffraction [74]. Especially the $g(OO)$ RDF misses the second peak in TIP3P. The improved TIP4P model yields this peak in better agreement with the experiment. However, the first peak is the most important one, because it determines the first solvation shell. For simulations of proteins in water the first

solvation shell is often a fair compromise. Introducing the TIP4P model in the Cornell et. al. force field would increase the computational effort because the position of the negative charge has to be additionally calculated in each step for each monomer.

The implementation of the water model in the CP-PAW framework is described in detail in [75]. In this work the RDF and the water dimer are calculated and compared with the experimental data. Also the constraints to keep the monomer rigid are discussed in detail.

4. Electronic structure calculations

Different methods exist to simulate systems on the atomic scale. While force fields, as discussed in Chapter 3, can describe quite large systems, they are unable to describe the behavior of chemical bonds during a reaction. To get access to the formation and breaking of the bonds one has to take the electrons of the system into account. The quantum mechanical description of such systems contains the many-particle wavefunction of the electrons and the nuclei. Because the masses of the nuclei are at least three orders of magnitude higher than the masses of the electrons, the equilibrium velocity of the electrons is also higher. This allows a decoupling of the electronic system from the system of the nuclei. The Born-Oppenheimer approximation separates the electronic and the nuclear wavefunctions. The electrons follow the motion of the atoms. In each atomic configuration the electrons can be considered to be in their ground state. Thus it is possible to consider only the energy of the electronic system at a given atomic geometry.

The many-particle Schrödinger equation for such a system is given by

$$\boldsymbol{H}\Psi = E\Psi \tag{4.0.1}$$

with the Hamiltonian:

$$\boldsymbol{H} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_e}\nabla_i^2 + \sum_{i=1}^{N} v_{ext}(\mathbf{r}_i) + \frac{e^2}{4\pi\varepsilon_0}\sum_{i<j}\frac{1}{r_{ij}} \tag{4.0.2}$$

Here, Ψ denotes the many-electron wave function, E is the total energy of the system and ∇_i^2 is the Laplacian with respect to the coordinates of electron i. v_{ext} is the external potential from the nuclei and r_{ij} is the distance between two electrons i and j. Using a variational principle the energy of the system is given by minimizing the functional:

$$E\left[\Psi\right] = \frac{\langle\Psi|\boldsymbol{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{4.0.3}$$

It is possible to solve Schrödinger's equation analytically for the hydrogen atom and other similar one-particle problems. But no solutions exist for many-body systems which are more complicated. However, some methods exist to solve these problems numerically but they are computationally very demanding. Therefore these methods are limited to small systems.

In the next section we discuss a method which allows to map the interacting many-particle system onto many non-interacting one-particle systems. This method is called density functional theory.

4.1. Density functional theory

4.1.1. Hohenberg-Kohn theorems

As shown before the ground state energy and wavefunction of any electronic system given by the Hamiltonian in equation (4.0.2) is determined by the minimization of the energy functional in equation (4.0.3). The external potential v_{ext} and the N electrons of the system define the Hamiltonian and thus all electronic properties of the system.

The electron density $n(\mathbf{r})$, following the first Hohenberg-Kohn theorem [76], determines the external potential $v_{ext}(\mathbf{r})$ and also the number of electrons N. If $n(\mathbf{r})$ is given, the Hamiltonian can be determined, thus leading to the many-particle ground state wave function Ψ and all other electronic properties.

This can be proved by a variational principle in the following way [77]: Consider a potential $v(\mathbf{r})$ that gives an electron density $n(\mathbf{r})$. Assuming there exists another potential $v'(\mathbf{r})$ which gives the same density, these potentials would give two Hamiltonians $\boldsymbol{H}$ and $\boldsymbol{H}'$. Both Hamiltonians would have the same ground state density but different wavefunctions Ψ and Ψ', with $\boldsymbol{H}\Psi = E_0\Psi$ and $\boldsymbol{H}'\Psi' = E_0'\Psi'$. The variational principle leads to:

$$\begin{aligned} E_0 &< \langle\Psi'|\boldsymbol{H}|\Psi'\rangle \\ &= \langle\Psi'|\boldsymbol{H}'|\Psi'\rangle + \langle\Psi'|\boldsymbol{H}-\boldsymbol{H}'|\Psi'\rangle \\ &= E_0' + \int n(\mathbf{r})\left[v(\mathbf{r})-v'(\mathbf{r})\right]\,\mathrm{d}^3r \end{aligned} \tag{4.1.1}$$

Similarly:

$$\begin{aligned} E_0' &< \langle\Psi|\boldsymbol{H}'|\Psi\rangle \\ &= \langle\Psi|\boldsymbol{H}|\Psi\rangle + \langle\Psi|\boldsymbol{H}'-\boldsymbol{H}|\Psi\rangle \\ &= E_0 - \int n(\mathbf{r})\left[v(\mathbf{r})-v'(\mathbf{r})\right]\,\mathrm{d}^3r \end{aligned} \tag{4.1.2}$$

Building the sum of equation (4.1.1) and (4.1.2), leads the contradiction $E_0 + E_0' < E_0 + E_0'$. It follows, that no two different $v(\mathbf{r})$ can yield the same $n(\mathbf{r})$, i.e. the density is a unique functional of the potential.

The energy functional with respect to the density is built by:

$$E[n(\mathbf{r})] = T[n(\mathbf{r})] + V_{en}[n(\mathbf{r})] + V_{ee}[n(\mathbf{r})] \tag{4.1.3}$$

$T[n(\mathbf{r})]$ is the kinetic energy. The potential is divided into two parts, V_{en} for the electron-nuclei interaction and V_{ee} for the electron-electron interaction. Equation (4.1.3) can also be written as:

$$E[n] = \int n(\mathbf{r})v_{ext}(\mathbf{r})\,\mathrm{d}^3r + F[n] \tag{4.1.4}$$

$F[n]$ is the sum of the kinetic energy and the electron-electron interaction.

$$F[n] = T[n] + V_{ee}[n] \tag{4.1.5}$$

It is independent of the external potential $v_{ext}(\mathbf{r})$ and thus valid for all systems. While the exact form of this functional is unknown, approximations have to be used. Some of them will be discussed later in section 4.1.3. The electron-electron interaction in equation (4.1.5) can be split into:

$$V_{ee}[n] = J[n] + \text{non-classical term} \tag{4.1.6}$$

$J[n]$ is the classical electron-electron repulsion:

$$J[n] = \frac{1}{2} \iint \frac{1}{r_{12}} n(\mathbf{r}_1) n(\mathbf{r}_2)\, \mathrm{d}\mathbf{r}_1 \mathrm{d}\mathbf{r}_2 \tag{4.1.7}$$

The non-classical term in equation (4.1.6) is a very important quantity. It is the major part of the exchange-correlation energy which will be discussed later in detail .

The second Hohenberg-Kohn theorem says, that for a given external potential $v_{ext}(\mathbf{r})$ the density $n(\mathbf{r})$ minimizes the true ground state energy E_0, which is an exactly defined functional of $n(\mathbf{r})$. Therefore it is

$$E_0 \leq E[\tilde{n}], \tag{4.1.8}$$

where $\tilde{n}$ is a trial density defined by $\int \mathrm{d}^3 r\, \tilde{n}(\mathbf{r}) = N$ and $E[\tilde{n}]$ is the energy functional of equation (4.1.4).

Hohenberg and Kohn only proved the existence of the energy functional. Later Kohn and Sham formulated the density functional theory in the way it is used today.

4.1.2. The Kohn-Sham Method

The method following the Hohenberg-Kohn theorems enables the determination of all electronic properties of a system defined by an electron density n. Unfortunately, the density functional $F[n]$ in (4.1.5) is unknown, as both parts of it, the kinetic energy $T[n]$ and the electron-electron interaction $V_{ee}[n]$ are unknown.

Kohn and Sham solved that problem in 1965 [78]. The idea behind Kohn-Sham DFT is the introduction of a virtual reference system of non-interacting electrons. The kinetic energy of the electrons can be written in terms of one-electron wave functions. The difference between the real system and the reference system is expressed by a correction term.

The kinetic energy of the ground state is given by

$$T = \sum_i^N n_i \langle \psi_i | -\tfrac{\hbar^2}{2m_e} \nabla^2 | \psi_i \rangle \tag{4.1.9}$$

where ψ_i are spin orbitals and n_i their occupation numbers. Following the Hohenberg-Kohn theorems T is a functional of the electron density given by:

$$n[\mathbf{r}] = \sum_i^N n_i \sum_s |\psi_i(\mathbf{r}, s)|^2 \tag{4.1.10}$$

The sums in these equations have an infinite number of terms in case of interacting systems. Kohn and Sham introduced a non-interacting reference system writing:

$$T_s[n] = \sum_i^N \langle\psi_i| - \tfrac{\hbar^2}{2m_e}\nabla^2|\psi_i\rangle \tag{4.1.11}$$

$$n[\mathbf{r}] = \sum_i^N \sum_s |\psi_i(\mathbf{r}, s)|^2 \tag{4.1.12}$$

Equation (4.1.11) and (4.1.12) are special cases of (4.1.9) and (4.1.10) with $n_i = 1$ for N orbitals and zero for the rest. Any nonnegative, continuous and normalized density n is N-representable and can be decomposed according to equation (4.1.12) [79]. The Hamiltonian of the non-interacting reference system is given by:

$$\boldsymbol{H}_s = \sum_i^N (-\tfrac{\hbar^2}{2m_e}\nabla_i^2) + \sum_i^N v_s(\mathbf{r}_i) \tag{4.1.13}$$

There are no electron-electron repulsions in the reference system. The wave function of the ground state is of one-determinant manner:

$$\Psi_s = \frac{1}{\sqrt{N!}} \det\left[\psi_1\psi_2\ldots\psi_N\right] \tag{4.1.14}$$

The states ψ_i are the N lowest eigenstates of the one-particle Hamiltonian which solve the following N equations:

$$\boldsymbol{h}_s\psi_i = \left[-\tfrac{\hbar^2}{2m_e}\nabla^2 + v_s(\mathbf{r})\right]\psi_i = \epsilon_i\psi_i \tag{4.1.15}$$

The kinetic energy $T_s[n]$ of this system is given by:

$$\begin{aligned} T_s[n] &= \langle\Psi_s| \sum_i^N \left(-\tfrac{\hbar^2}{2m_e}\nabla_i^2\right) |\Psi_s\rangle \\ &= \sum_{i=1}^N \langle\psi_i| - \tfrac{\hbar^2}{2m_e}\nabla^2|\psi_i\rangle \end{aligned} \tag{4.1.16}$$

It can be calculated exactly and is close to the real kinetic energy $T[n]$. Equation (4.1.5) can be rewritten to separate $T_s[n]$ as the kinetic energy component:

$$F[n] = T_s[n] + J[n] + E_{xc}[n] \tag{4.1.17}$$

where

$$E_{xc}[n] = T[n] - T_s[n] + V_{ee}[n] - J[n] \tag{4.1.18}$$

$E_{xc}[n]$ is the exchange-correlation energy. It contains the difference between $T[n]$ and $T_s[n]$ and the non-classical part of $V_{ee}[n]$. This non-classical part leads

to the so-called exchange-correlation hole which is caused by the Pauli-repulsion of the electrons.

A non-interacting reference system of N equations

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + v_{eff}(\mathbf{r})\right)\psi_i = \epsilon_i\psi_i \tag{4.1.19}$$

is solved by the one-particle states ψ_i. Together with the density

$$n(\mathbf{r}) = \sum_i^N \sum_s |\psi_i(\mathbf{r}, s)|^2 \tag{4.1.20}$$

and the exchange-correlation potential

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \tag{4.1.21}$$

the Kohn-Sham effective potential is defined by:

$$\begin{aligned} v_{eff}(\mathbf{r}) &= v_{ext}(\mathbf{r}) + \frac{\delta J[n]}{\delta n(\mathbf{r})} + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})} \\ &= v_{ext}(\mathbf{r}) + \frac{1}{4\pi\varepsilon_0}\int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}' + v_{xc}(\mathbf{r}) \end{aligned} \tag{4.1.22}$$

The equations (4.1.19) - (4.1.22) are the so-called Kohn-Sham equations. The effective potential in equation (4.1.22) depends on the density $n(\mathbf{r})$. Therefore, the other Kohn-Sham equations have to be solved self-consistently. This is done by guessing an initial density $n(\mathbf{r})$ and thus constructing the effective potential. In the next step a new density can be calculated from equations (4.1.19) and (4.1.20).

The total energy functional is given by:

$$E[n] = T_s[n] + J[n] + \int v_{ext}(\mathbf{r})n(\mathbf{r})\,\mathrm{d}^3r + E_{xc}[n] \tag{4.1.23}$$

All terms of the total energy functional can be calculated exactly, except the exchange-correlation functional $E_{xc}[n]$. Some approaches of its approximation are presented in the following section.

4.1.3. Approximations for the exchange and correlation

The construction of accurate approximation techniques for the exchange-correlation functional in equation (4.1.23) is still subject of modern research. In the following two major methods are presented.

The local density approximation

The simplest approximation of $E_{xc}[n]$ is the so-called local density approximation (LDA). This model is described by a system of a homogenous electron gas. This is a uniform distribution of electrons, which move on a posi-

tively charged background to ensure electro-neutrality. The electron gas should have the same density as the real system. $E_{xc}[n]$ is constructed by taking the exchange-correlation energy per particle to be equal to the one of the uniform electron gas of the same density, $\epsilon_{xc}[n]$, and weighting it with the density:

$$E_{xc}^{LDA}[n] = \int \epsilon_{xc}(n) n(\mathbf{r}) \, \mathrm{d}^3 r \tag{4.1.24}$$

The exchange-correlation energy of the electron gas can be divided into the contribution of the exchange part and the correlation part:

$$\epsilon_{xc}(n) = \epsilon_x(n) + \epsilon_c(n) \tag{4.1.25}$$

The exchange contribution is known analytically for the uniform electron gas. It can be calculated exactly from the Dirac exchange-energy functional:

$$\epsilon_x^{LDA}(n) = -\frac{e^2}{4\pi\varepsilon_0}\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n(\mathbf{r})^{1/3} \tag{4.1.26}$$

Unfortunately, no explicit expression for the correlation part exists, although its high-density and its low-density limits are known. $\epsilon_c(n)$ has been parameterized by fitting an analytical form to the very accurate quantum Monte Carlo results for the homogenous electron gas. The correlation energies are an order of magnitude smaller than the corresponding exchange energies, thus contributing only a small part to $\epsilon_{xc}(n)$ [77]. All QM simulations during the work on this thesis were performed using the parameterization of Perdew and Wang [80] for the local correlation energy.

Although LDA cannot be expected to yield any reasonable results for real inhomogeneous systems, it turned out that LDA performs remarkably well. This helped to increase the acceptance of DFT over the first years. Structural parameters are accurate to about 1%, bond energies are generally overestimated by LDA [81, 82].

Generalized gradient approximations

As mentioned above the accuracy for energies by LDA is not generally sufficient in chemical applications. So far only the information of the local density $n(\mathbf{r})$ itself has been used in the functional. The next logical step is to include also the gradient of the density $\nabla n(\mathbf{r})$. The gradient is used to take the inhomogeneity of the electron density into account:

$$E_{xc}^{GGA}[n] = \int n(\mathbf{r}) \epsilon_{xc}^{GGA}(n, |\nabla n|) \, \mathrm{d}^3 r \tag{4.1.27}$$

This method is called generalized gradient approximation (GGA). Typically these functions are constructed by adding gradient corrections to LDA functionals.

GGA functionals are often parameterized by fitting experimental data. Other methods use physical boundary conditions to construct the functionals [77]. In

all calculations done during the work on this thesis the parameter-free GGA by Perdew, Burke and Ernzerhof (PBE) [83] was used. This functional has not been fitted to experimental data, thus it is valid for a wide range of systems. It reproduces experimental data significantly better than LDA. The correlation energy in the PBE functional is given by:

$$\begin{aligned} \epsilon_C^{PBE} &= \epsilon_c^{LDA} - \gamma\phi^3 \cdot \ln\left[1 + \frac{1}{\chi s^2/\phi^2 + (\chi s^2/\phi^2)^2}\right] \\ &\gamma = \frac{1-\ln 2}{\pi^2} \approx 0.031091 \\ &\phi(\zeta) = \tfrac{1}{2}\left[(1+\zeta)^{2/3} + (1-\zeta)^{2/3}\right], \quad \zeta = \frac{n_\uparrow - n_\downarrow}{n} \\ &\chi \approx 0.72161, \quad s = \frac{|\nabla n|}{2n(3\pi^2 n)^{1/3}} \end{aligned} \tag{4.1.28}$$

The exchange energy is constructed by:

$$\begin{aligned} \epsilon_x^{PBE} &= \epsilon_x^{LDA} \cdot F_x(s) \\ F_x(s) &= 1 + \kappa - \frac{\kappa}{1 - \mu s^2/\kappa} \\ &\kappa = 0.804, \quad \mu \approx 0.21951 \end{aligned} \tag{4.1.29}$$

The PBE gradient correction has been chosen because it is known for its general applicability and it is widely accepted and known to give accurate results for a wide range of systems.

4.2. The projector augmented wave method

The electronic structure method of choice in all QM calculations within this thesis is the projector augmented wave method (PAW) developed by Blöchl. It is implemented in the CP-PAW program package. In this section I will describe the ideas behind PAW and how it works. I will follow the derivation of the method as it is described in [84]. The complete derivation can be found in the original paper [85].

4.2.1. The basic ideas

The electronic wave function of a system behaves different depending on the position in the system. On the one hand is the atomic region near the nucleus on the other hand is the interstitial region, the region of chemical bonds. In the first the kinetic energy of the electrons is large. This results in rapid oscillations of the wavefunction and therefore fine grids are needed to describe it. The shape of the wavefunction is hardly influenced by changes in the chemical environment. Therefore a small basis set is able to describe the wavefunction in the atomic region properly. In the latter the kinetic energy is small and the wavefunction is smooth. This is the region of chemical bonds and the wavefunction is strongly influenced by the environment. A large basis set - nearly a complete one - is required to describe the wavefunction in this region.

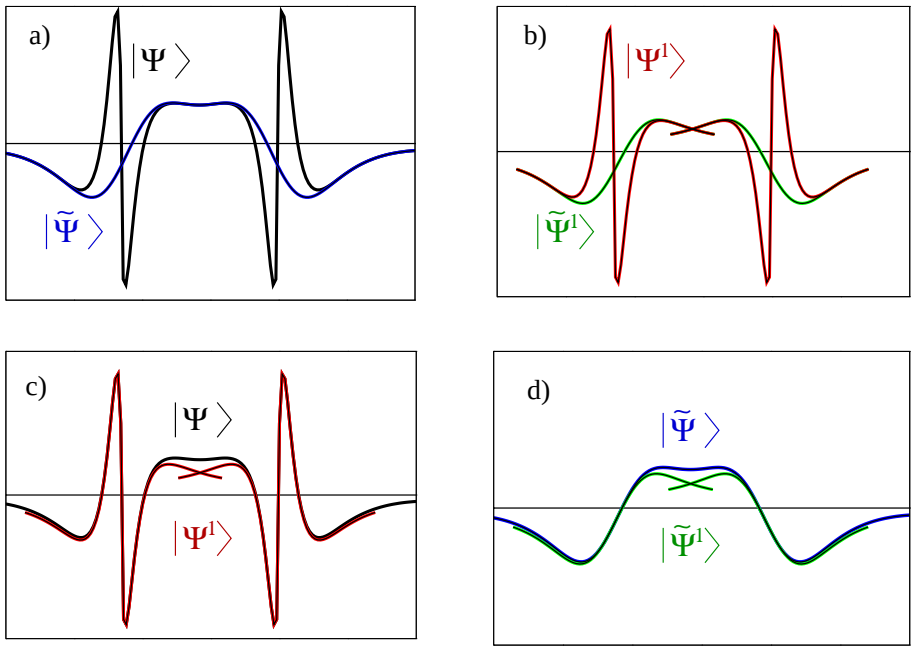


Figure 4.1: The wave functions in PAW: a) the real and the auxiliary wave function; b) the one-center expansion of the true and of the auxiliary wave function; c) the real wave function and its one-center expansion; d) the auxiliary wave function and its one-center expansion. The pictures are taken with kind permission from [84].

In the region of chemical bonds the wavefunction is best described by plane waves which can describe the smooth character in the bonding region and which are effectively treatable on computers. In the atomic region the plane waves have to be augmented with some functions in order to describe the spikes and nodes.

Before discussing the mathematical basis of PAW, the basic ideas behind PAW shall be described. Consider the p-σ-orbital of a Cl_2 molecule. Figure 4.1a shows the physical wave function $|\Psi\rangle$. The blue line shows the auxiliary wave function $|\tilde{\Psi}\rangle$ which is used in the PAW method. As one can easily see, $|\tilde{\Psi}\rangle$ matches the physical wave function in the interstitial region. However it cannot describe the strong oscillation at the atomic sites. This error is corrected by the one-center expansion $|\Psi^1\rangle$ within the atomic region, which is shown in Figure 4.1c. This one-center expansion of the real wave function is derived from the one-center expansion of the auxiliary wave function $|\tilde{\Psi}^1\rangle$(Fig. 4.1b). The real wave function is the sum of the auxiliary wave function and the difference of the one-center expansions.

$$|\Psi\rangle = |\tilde{\Psi}\rangle + \left(|\Psi^1\rangle - |\tilde{\Psi}^1\rangle\right) \tag{4.2.1}$$

In the next section the mathematical basis of the PAW method, the transformation theory, will be described. The goal is to construct a transformation which maps the true wave function onto auxiliary wave functions.

4.2.2. The transformation operator

Consider a wave function $|\Psi\rangle$. Its complete nodal structure shall be mapped onto an auxiliary wave function $|\tilde{\Psi}\rangle$, which has a smooth character. Its smoothness allows it to be easily expanded into a plane wave basis set, which converges quite fast. The transformation from the auxiliary to the physical wave function is done by the transformation operator $\boldsymbol{T}$:

$$|\Psi\rangle = \boldsymbol{T}|\tilde{\Psi}\rangle \tag{4.2.2}$$

By this transformation the Schrödinger equation

$$\boldsymbol{H}|\Psi\rangle = |\Psi\rangle E \tag{4.2.3}$$

with the Kohn-Sham Hamilton operator $\boldsymbol{H}$ becomes:

$$\boldsymbol{T}^\dagger \boldsymbol{H}\boldsymbol{T}|\tilde{\Psi}\rangle = \boldsymbol{T}^\dagger \boldsymbol{T}|\tilde{\Psi}\rangle E \tag{4.2.4}$$

The next step is to define the transformation operator $\boldsymbol{T}$.

The auxiliary wave function has to be transformed by $\boldsymbol{T}$, so that the smooth wave function has the correct nodal structure in each atomic region. Thus, the transformation is written as identity plus a sum of atomic contributions $\boldsymbol{S}_R$:

$$\boldsymbol{T} = 1 + \sum_R \boldsymbol{S}_R \tag{4.2.5}$$

R labels the atomic site. The operator $\boldsymbol{S}_R$ in equation (4.2.5) adds the difference between the true and the auxiliary wave function for each atom. It is constructed by a basis set of partial waves $|\phi_i\rangle$. These partial waves are solutions of the Schrödinger equation of isolated atoms. Near the nucleus the valence wave functions are expressed by a superposition of the partial waves:

$$\Psi(\mathbf{r}) = \sum_{i \in R} \phi_i(\mathbf{r}) c_i \tag{4.2.6}$$

for $|\mathbf{r} - \mathbf{R}_R| < r_{c,R}$. $\mathbf{R}_R$ is the position of the nucleus of site R. The index i describes all partial waves which belong to R, containing the angular momentum indices (l, m). $r_{c,R}$ is the cutoff radius, which ensures the normalizability of the partial waves, hence they do not necessarily belong to bound states. The set of partial waves includes only valence states, which are orthogonal to the core states. The frozen core approximation is used, i.e. the core electrons are the same as the corresponding electrons of an isolated atom.

An auxiliary partial wave $|\tilde{\phi}_i\rangle$ belongs to each partial wave $|\phi_i\rangle$. This defines the local contribution $\boldsymbol{S}_R$ of the transformation operator as:

$$|\phi_i\rangle = (1 + \boldsymbol{S}_R)|\tilde{\phi}_i\rangle \tag{4.2.7}$$

$$\boldsymbol{S}_R|\tilde{\phi}_i\rangle = |\phi_i\rangle - |\tilde{\phi}_i\rangle \tag{4.2.8}$$

The transformation operator changes the wave function only locally. This makes it necessary that the partial waves and their auxiliary wave functions become

identical beyond the cutoff radius, i.e. $\phi_i(\mathbf{r}) = \tilde{\phi}_i(\mathbf{r})$ for all $|\mathbf{r} - \mathbf{R}| < r_{c,R}$.

Applying the transformation operator to an auxiliary wave function, $\tilde{\Psi}(\mathbf{r})$ has to be expanded into auxiliary partial waves:

$$\tilde{\Psi}(\mathbf{r}) = \sum_i \tilde{\phi}_i(r) \cdot c_i \tag{4.2.9}$$

At this point the projector function $|\tilde{p}_i\rangle$ is introduced. It defines the coefficients c_i:

$$c_i = \int \tilde{p}_i^*(\mathbf{r}) \tilde{\Psi}_i(\mathbf{r}) \, \mathrm{d}^3 r = \langle \tilde{p}_i | \tilde{\Psi}_i \rangle \tag{4.2.10}$$

This leads to:

$$\tilde{\Psi}(\mathbf{r}) = \sum_{i \in R} \tilde{\phi}_i(\mathbf{r}) \langle \tilde{p}_i | \tilde{\Psi}_i \rangle \tag{4.2.11}$$

for all $|\mathbf{r} - \mathbf{R}| < r_{c,R}$. The projector functions probe the character of the auxiliary wave function in the atomic region. It is obvious that $\sum_i |\tilde{\phi}_i\rangle\langle\tilde{\Psi}_i| = 1$ if $\mathbf{r}$ lies inside the cutoff radius. Therefore, the one-center expansion $\sum_i |\tilde{\phi}_i\rangle\langle\tilde{p}|\tilde{\Psi}\rangle$ of an auxiliary wave function is identical to the auxiliary wave function $|\tilde{\Psi}\rangle$ itself. This implies:

$$\langle \tilde{p}_i | \tilde{\Psi}_i \rangle = \delta_{ij} \tag{4.2.12}$$

where δ_{ij} is the Kronecker symbol.

Together with equations (4.2.8) and (4.2.11) the local contribution $\boldsymbol{S}_R$ to the transformation operator can be applied to any auxiliary wave function:

$$\begin{aligned} \boldsymbol{S}_R |\tilde{\Psi}\rangle &= \sum_{i \in R} \boldsymbol{S}_R |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \tilde{\Psi}\rangle \\ &= \sum_{i \in R} \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle \right) \langle \tilde{p}_i | \tilde{\Psi}\rangle \end{aligned} \tag{4.2.13}$$

and the transformation operator $\boldsymbol{T}$ can be written as:

$$\boldsymbol{T} = 1 + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle \right) \langle \tilde{p}_i | \tag{4.2.14}$$

Bringing all together, the true wave function can be expressed as:

$$\begin{aligned} |\Psi\rangle &= |\tilde{\Psi}\rangle + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle \right) \langle \tilde{p}_i | \tilde{\Psi}\rangle \\ &= |\tilde{\Psi}\rangle + \left(\sum_{i \in R} |\phi_i\rangle \langle \tilde{p}_i | \tilde{\Psi}\rangle - \sum_{i \in R} |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \tilde{\Psi}\rangle \right) \\ &= |\tilde{\Psi}\rangle + \sum_R \left(|\Psi_R^1\rangle - |\tilde{\Psi}_R^1\rangle \right) \end{aligned} \tag{4.2.15}$$

Equation (4.2.15) can be understood as follows: The auxiliary wave function is identical to the true wave function in a region which lies far away from the atom centers. The partial waves are identical with the partial wave expansion of the auxiliary wave function. In the region around the atom centers the true wave function is built up from partial waves which contain the nodal structure. The auxiliary wave function and its partial wave expansion are equal.

For physical quantities of interest the expectation value of an operator $\boldsymbol{A}$ has to be evaluated. This can either be done with the reconstructed true wave function or with the auxiliary wave functions:

$$\langle \boldsymbol{A} \rangle = \sum_n f_n \langle \Psi_n | \boldsymbol{A} | \Psi_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \boldsymbol{A} | \phi_n^c \rangle \tag{4.2.16}$$

$$= \sum_n f_n \langle \tilde{\Psi}_n | \boldsymbol{T}^\dagger \boldsymbol{A} \boldsymbol{T} | \tilde{\Psi}_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \boldsymbol{A} | \phi_n^c \rangle \tag{4.2.17}$$

The first sum runs over the valence states whose occupations are denoted by f_n. The second sum runs over N_c core states $|\phi_n^c\rangle$. The evaluation of an expectation value can be expressed by the same decomposition shown in equation (4.2.15):

$$\begin{aligned}
\langle \Psi | \boldsymbol{A} | \Psi \rangle &= \langle \tilde{\Psi} + \sum_R \left(\Psi_R^1 - \tilde{\Psi}_R^1 \right) | \boldsymbol{A} | \tilde{\Psi} + \sum_{R'} \left(\Psi_{R'}^1 - \tilde{\Psi}_{R'}^1 \right) \rangle \\
&= \underbrace{\langle \tilde{\Psi} | \boldsymbol{A} | \tilde{\Psi} \rangle + \sum_R \left(\langle \Psi_R^1 | \boldsymbol{A} | \Psi_R^1 \rangle - \langle \tilde{\Psi}_R^1 | \boldsymbol{A} | \tilde{\Psi}_R^1 \rangle \right)}_{\text{part 1}} \\
&\quad + \underbrace{\sum_R \left(\langle \Psi_R^1 - \tilde{\Psi}_R^1 | \boldsymbol{A} | \tilde{\Psi} - \tilde{\Psi}_R^1 \rangle + \langle \tilde{\Psi} - \tilde{\Psi}_R^1 | \boldsymbol{A} | \Psi_R^1 - \tilde{\Psi}_R^1 \rangle \right)}_{\text{part 2}} \\
&\quad + \underbrace{\sum_{R \neq R'} \langle \Psi_R^1 - \tilde{\Psi}_R^1 | \boldsymbol{A} | \Psi_{R'}^1 - \tilde{\Psi}_{R'}^1 \rangle}_{\text{part 3}}
\end{aligned} \tag{4.2.18}$$

The expression in equation (4.2.18) has to be discussed in more detail.

The first term in part two contains the function $\Psi_R^1 - \tilde{\Psi}_R^1$. This function vanishes beyond the augmentation region because the partial waves are pairwise identical in this region as discussed earlier. The second function in this term is $\tilde{\Psi} - \tilde{\Psi}_R^1$. Inside the augmentation region this function vanishes, too. It follows that the first term in part two has in every region of space a zero component, thus the whole term becomes zero. The second term in part two contains also the function $\Psi_R^1 - \tilde{\Psi}_R^1$. Both parts of this function will not become simultaneously non-zero in the same region. Following this argumentation part two and three of equation (4.2.18) vanish if they are evaluated for local operators. Examples of local operates are the real space projection operator $|r\rangle\langle r|$ which yields the electron density or the kinetic energy operator $-\frac{\hbar^2}{2m_e}\nabla^2$. In the case of non-local operators these parts have to be evaluated explicitly.

Introducing the one-center density matrix

$$D_{i,j} = \sum_n f_n \langle \tilde{\Psi}_n | \tilde{p}_j \rangle \langle \tilde{p}_i | \tilde{\Psi}_n \rangle \tag{4.2.19}$$

the expectation value $\langle \boldsymbol{A} \rangle$ in equation (4.2.16) can be expressed as:

$$\begin{aligned}
\langle \boldsymbol{A} \rangle &= \sum_n f_n \left(\langle \tilde{\Psi}_n | \boldsymbol{A} | \tilde{\Psi}_n \rangle + \langle \Psi_n^1 | \boldsymbol{A} | \Psi_n^1 \rangle - \langle \tilde{\Psi}_n^1 | \boldsymbol{A} | \tilde{\Psi}_n^1 \rangle \right) + \sum_{n=1}^{N_c} \langle \phi_n^c | \boldsymbol{A} | \phi_n^c \rangle \\
&= \sum_n f_n \langle \tilde{\Psi}_n | \boldsymbol{A} | \tilde{\Psi}_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \boldsymbol{A} | \phi_n^c \rangle \\
&\quad + \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \phi_j | \boldsymbol{A} | \phi_i \rangle + \sum_{n \in R}^{N_c, R} \langle \phi_n^c | \boldsymbol{A} | \phi_n^c \rangle \right) \\
&\quad - \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_j | \boldsymbol{A} | \tilde{\phi}_i \rangle + \sum_{n \in R}^{N_c, R} \langle \tilde{\phi}_n^c | \boldsymbol{A} | \tilde{\phi}_n^c \rangle \right)
\end{aligned} \tag{4.2.20}$$

$|\tilde{\phi}_n^c\rangle$ are the auxiliary core states. They have been introduced into the equation to allow the incorporation of the tails of the core wave function into the plane wave part. This ensures that the integration of partial wave contributions cancel strictly beyond r_c.

Finally, the electron density can be expressed as

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_R \left(n_R^1(\mathbf{r}) - \tilde{n}_R^1(\mathbf{r}) \right) \tag{4.2.21}$$

within the PAW frame work. The three parts of the density are:

$$\begin{aligned}
\tilde{n}(\mathbf{r}) &= \sum_n f_n \tilde{\Psi}_n^*(\mathbf{r}) \tilde{\Psi}_n(\mathbf{r}) + \tilde{n}_c \\
n_R^1(\mathbf{r}) &= \sum_{i,j \in R} D_{i,j} \phi_j^*(\mathbf{r}) \phi_i(\mathbf{r}) + n_{c,R} \\
\tilde{n}_R^1(\mathbf{r}) &= \sum_{i,j \in R} D_{i,j} \tilde{\phi}_j^*(\mathbf{r}) \tilde{\phi}_i(\mathbf{r}) + \tilde{n}_{c,R}
\end{aligned} \tag{4.2.22}$$

where $n_{c,R}$ is the core density of the corresponding atom and $\tilde{n}_c, R$ is the auxiliary core density.

The total energy can also be divided into three parts:

$$E\left([\tilde{\Psi}_n], R_i \right) = \tilde{E} + \sum_R \left(E_R^1 - \tilde{E}_R^1 \right) \tag{4.2.23}$$

$\tilde{E}$ is the plane wave part. It is evaluated on equispaced grids in real and recip-

rocal space. It is given by:

$$\begin{aligned}\tilde{E} &= \sum_n \langle\tilde{\Psi}_n|\frac{-\hbar^2}{2m_e}\nabla^2|\tilde{\Psi}_n\rangle \\ &+\frac{e^2}{8\pi\varepsilon_0}\int \mathrm{d}^3r \int \mathrm{d}^3r' \frac{\left(\tilde{n}(\mathbf{r})+\tilde{Z}(\mathbf{r})\right)\left(\tilde{n}(\mathbf{r}')+\tilde{Z}(\mathbf{r}')\right)}{|\mathbf{r}-\mathbf{r}'|} \\ &+\int \mathrm{d}^3r\, \tilde{n}(\mathbf{r})\epsilon_{xc}(\mathbf{r},[\tilde{n}]) + \int \mathrm{d}^3r\, \bar{v}(\mathbf{r})\tilde{n}(\mathbf{r}) \end{aligned} \tag{4.2.24}$$

$\tilde{Z}(\mathbf{r}) = \sum_R \tilde{Z}_R(\mathbf{r})$ is an angular momentum dependent core-like compensation density. It is a sum of Gauss functions, which have an analytical Fourier transform. It ensures that the augmentation charge densities

$$n^1_R(\mathbf{r}) + Z_R(\mathbf{r}) - \tilde{n}^1_R(\mathbf{r}) - \tilde{Z}_R(\mathbf{r})$$

have vanishing electrostatic multi-pole moments for each atomic site. Therefore, the sum over all one-center contributions from one atom has no contribution to the electrostatic potential outside the atomic region. The electrostatic interaction of the one-center parts between different sites vanishes.

The remaining terms in Equation (4.2.23) are given as:

$$\begin{aligned}E^1_R &= \sum_{i,j\in R} D_{i,j}\langle\phi_j|\frac{-\hbar^2}{2m_e}\nabla^2|\phi_i\rangle + \sum_{n\in R}^{N_{c,R}}\langle\phi^c_n|\frac{-\hbar^2}{2m_e}\nabla^2|\phi^c_n\rangle \\ &+\frac{e^2}{8\pi\varepsilon_0}\int \mathrm{d}^3r \int \mathrm{d}^3r' \frac{\left(n^1(\mathbf{r})+Z(\mathbf{r})\right)\left(n^1(\mathbf{r}')+Z(\mathbf{r}')\right)}{|\mathbf{r}-\mathbf{r}'|} \\ &+\int \mathrm{d}^3r\, n^1(\mathbf{r})\epsilon_{xc}(\mathbf{r},[n^1]) \end{aligned} \tag{4.2.25}$$

$$\begin{aligned}\tilde{E}^1_R &= \sum_{i,j\in R} D_{i,j}\langle\tilde{\phi}_j|\frac{-\hbar^2}{2m_e}\nabla^2|\tilde{\phi}_i\rangle + \sum_{n\in R}^{N_{c,R}}\langle\phi^c_n|\frac{-\hbar^2}{2m_e}\nabla^2|\phi^c_n\rangle \\ &+\frac{e^2}{8\pi\varepsilon_0}\int \mathrm{d}^3r \int \mathrm{d}^3r' \frac{\left(\tilde{n}^1(\mathbf{r})+\tilde{Z}(\mathbf{r})\right)\left(\tilde{n}^1(\mathbf{r}')+\tilde{Z}(\mathbf{r}')\right)}{|\mathbf{r}-\mathbf{r}'|} \\ &+\int \mathrm{d}^3r\, \tilde{n}^1(\mathbf{r})\epsilon_{xc}(\mathbf{r},[\tilde{n}^1]) + \int \mathrm{d}^3r\bar{v}(\mathbf{r})\tilde{n}^1(\mathbf{r}) \end{aligned} \tag{4.2.26}$$

Starting from the energy functional given in Equation (4.2.23) other quantities can be derived: The Hamiltonian can be constructed by deriving the total energy with respect to the wave functions. The forces are partial derivatives of the total energy with respect to atomic position and the potential is the derivative of the potential energy with respect to the density.

4.2.3. Approximations in the PAW method

The PAW method is an exact formulation of the Kohn-Sham equations. Nevertheless some approximations have to be done implementing the PAW method in

practice. The following approximations have been included in the total energy functional of the PAW method:

- The frozen core approximation. Only valence electrons are taken into account. Electrons of lower lying shells are part of the core.
- The plane wave expansion for the auxiliary wave functions must be complete. The plane wave expansion is controlled easily by increasing the plane wave cutoff. The cutoff is defined by $E_{PW} = \frac{1}{2}\hbar^2 G_{max}^2$. In all calculations in this work a cutoff of 30 Ry has been used.
- The partial wave expansion must be converged. Typically one or two partial waves are used per angular momentum (l, m) and site. For the partial wave expansion changes the total energy functional, it is not variational.

All errors in the PAW method can be systematically controlled. No transferability errors occur. The PAW method provides access to the full charge and spin density, which is relevant for hyperfine parameters. Hyperfine parameters are sensitive probes of the electron density near the nucleus. In many situations they are the only information available that allows to deduce atomic structure and chemical environment of an atom.

4.3. Molecular Dynamics

Simulating molecules at finite temperature requires the description of the atomic motion. This is called molecular dynamics. Molecular dynamics can be used to analyze the frequencies in a molecule, chemical reactions (using constraint motions if the reaction barrier is too high) or finding the ground state of a system by applying friction to the atomic motion. Each atomic motion is driven by the forces acting on the atoms. These forces can be calculated by the total energy of the system:

$$F_i = \frac{\partial E}{\partial R_i} \tag{4.3.1}$$

The total energy of the system can be obtained from the electronic structure based on DFT and the PAW method.

4.3.1. The equations of motion

The motion of the atoms in a system can be described by Newton's equations of motion:

$$F_i = M_i \ddot{R}_i \tag{4.3.2}$$

The index i runs over all atoms and over all degrees of freedom per atom. To calculate the ground state of a given system, energy has to be damped out by applying friction to the equation of motion. Equation (4.3.2) becomes:

$$F_i = M_i \ddot{R}_i + M_i \alpha \dot{R}_i \tag{4.3.3}$$

To solve the equations of motion in a computer simulation, they have to be discretized in time steps Δt. $\dot{R}$ and $\ddot{R}$ will be replaced by the differential quotients:

$$\dot{R}_i \approx \frac{R_i(t+\Delta t) - R_i(t-\Delta t)}{2\Delta t} \tag{4.3.4}$$

$$\ddot{R}_i \approx \frac{R_i(t+\Delta t) - 2R_i(t) + R_i(t-\Delta t)}{(\Delta t)^2} \tag{4.3.5}$$

This method, which is also used in the CP-PAW program package, is called Verlet algorithm [86]. While $R_i(t)$ holds the atomic positions in the actual time step, $R_i(t-\Delta t)$ describes the atomic positions in the last time step, $R_i(t+\Delta t)$ the positions in the following time step. The combination of equation (4.3.5) with the equation of motion (4.3.3) leads to a discretized damped equation of motion:

$$R_i(t+\Delta t) = \frac{1}{1+a}\left(2R_i(t) - (1-a)R_i(t-\Delta t) + \frac{F_i(\Delta t)^2}{M_i}\right) \tag{4.3.6}$$

with $a = \alpha\Delta t/2$. The time step Δt has to be chosen, that the desired accuracy is reached but the number of iterations is minimized. The Verlet algorithm is limited by the vibration modes of the simulated system. For time steps which are larger than about 30% of the smallest occurring modes, the algorithm diverges.

4.3.2. Car-Parrinello molecular dynamics

Describing the dynamics of a system from first principles is called *ab initio* molecular dynamics (AIMD). This means that no prefabricated potentials are taken into account (neither empirical nor from independent electronic structure calculations). The forces determining the dynamics of the system have thus to be extracted ab initio from the instantaneous state of the system. That is, the electronic structure problem is solved for a given geometry of the system. This leads to the forces acting on the system. In the next step the system is propagated according these forces which yields a new geometry. The wave functions have to be optimized again, which leads to an update of the forces, and so on. This method requires a large computational effort and is not very practicable.

Because of the mass difference between electrons and atoms the Born-Oppenheimer approximation is valid, which allows the separation of the electron and nuclear motion. This allows to solve the time-independent Schrödinger equation for the electrons at a fixed configuration of the atoms. Car and Parrinello developed a formalism which allows the simultaneous simulation of the motion of electrons and nuclei [87], the so-called Car-Parrinello molecular dynamics. They describe the electronic motion by a fictitious motion keeping the electrons in the ground state. An extended Lagrangian is constructed, which combines the nuclei and wave functions:

$$\mathcal{L} = \tfrac{1}{2}\sum_I M_I \dot{\mathbf{R}}_I^2 + \sum_i m_\psi f_i \langle\dot{\psi}|\dot{\psi}\rangle - E\left[\{\mathbf{R}_I\},\{\psi_i\}\right] + \sum_{i,j}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right)\Lambda_{ji}$$

(4.3.7)

The first term describes the classical kinetic energies of the nuclei I at position $\mathbf{R}_I$ with masses M_I. The second term introduces a fictitious kinetic energy of the electrons with a fictitious mass m_ψ. The potential energy for the nuclear and wave functions degrees of freedom comes from the energy functional in the third term. It depends on the atomic positions and the one-electron wave functions. The wave functions have to be orthonormal, which is ensured by the constraints in the fourth term, using the method of Lagrange multipliers.

The Lagrangian in equation (4.3.7) generates Newtonian equations of motion for the electrons and the nuclei. The Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} - \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I} = 0 \tag{4.3.8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\delta \mathcal{L}}{\delta \langle \dot{\psi}_i|} - \frac{\delta \mathcal{L}}{\delta \langle \psi|_i} = 0 \tag{4.3.9}$$

yield the coupled equations of motion for both:

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E[\{\mathbf{R}_I\},\{\psi_i\}]}{\partial \mathbf{R}_I} = -\nabla_I E[\{\mathbf{R}_I\},\{\psi_i\}] = \mathbf{F}_I \tag{4.3.10}$$

$$\begin{aligned} m_\psi |\ddot{\psi}_i\rangle &= -\frac{\delta E[\{\mathbf{R}_I\},\{\psi_i\}]}{\delta \langle \psi_i|} + \tfrac{1}{f_i}\sum_j |\psi_j\rangle \Lambda_{ji} \\ &= -\boldsymbol{H}|\psi_i\rangle + \tfrac{1}{f_i}\sum_j |\psi_j\rangle \Lambda_{ji} \end{aligned} \tag{4.3.11}$$

The Hamiltonian in equation (4.3.11) is the Kohn-Sham Hamiltonian which yields the ground state energy of the electronic system. The atoms are moved according to the forces obtained from (4.3.10). As the nuclei move, the wave functions are propagated simultaneously, adapting to the new nuclear configuration without requiring any self-consistency cycles. Most simulations start with a fixed atom geometry. The initial random wave functions are optimized for that special geometry. In the next step, the nuclei are released and propagated according the forces. The wave functions stay in, or close to, the ground state during the time evolution. They move on, or close to, the Born-Oppenheimer surface. Figure 4.2 illustrates the movement of the system on the Born-Oppenheimer surface.

Finally, it has to be emphasized that the dynamics of the wave functions is completely fictitious and has nothing to do with the real physics of the wave functions.

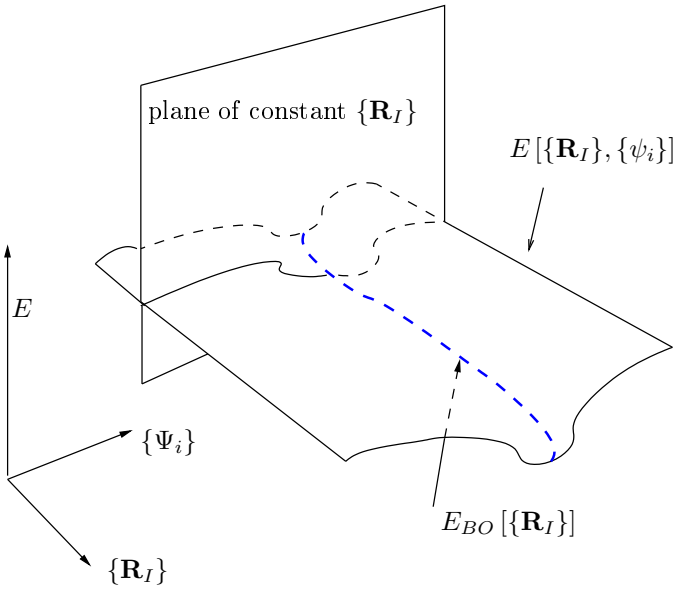


Figure 4.2: The Born-Oppenheimer surface

Part II.

Applications

5. The Implentation of the AMBER Force Field in the CP-PAW Program Package

QM/MM simulations of biomolecules with the reaction center in the QM part and the rest of the biomolecule in the MM part require a force field which is optimized for such systems. The CP-PAW program package contains already the universal force field (UFF). The advantage of UFF is that it is able to treat all elements of the periodic table. But it fails in describing hydrogen bonds or special geometries of peptide bonds.

The Cornell et. al. force field (also called AMBER force field) has been developed for biological systems. In the following sections the implementation of the AMBER force field into the CP-PAW program package is described in detail. This implementation has been guided by the idea to use the existing code structures and to expand them in a way which makes it possible to easily include other force fields.

5.1. AMBER and CP-PAW

5.1.1. Input format for the MM system

The QM/MM implementation in CP-PAW shall be used to simulate large systems on the MM level with small inner subsystems which are treated on the QM level. CP-PAW provides actually two force fields: UFF and AMBER. As mentioned above UFF has been the first implementation. The MM system is defined in the structure file which is written in an XML-like format. Examples for the input format can be found in appendix A and at the end of this chapter.

Tests with larger systems (several thousand atoms) show that the calculations become very slow. The problem is that the way the input data is handled is computationally very expensive. On the other hand most biological structures like proteins come in the well-defined PDB format. PDB files are formatted files. Data lines begin with a keyword and this decides in which column which information is written. Choosing PDB files as input format for QM/MM calculations has also the advantage that the AMBER force field uses the same name convention. So it is quite straightforward to use the PDB files. An example of a PDB file can be found at the end of this chapter.

To make the input data even more comfortable the PDB format has been extended by three data columns. Within these columns the coupling between QM and MM atoms is defined. This will be explained in detail in the following sections. Table D.1 in the appendix D.1 shows the details of the input format

which I use in CP-PAW. In the same section one can find the FORTRAN format string which can be used to read and write the PDB files. Using the force field module, this format string is available in other program parts. The details of the read-in procedure are described in section 5.1.3.

The heart of the AMBER force field consists of two files, the parameter and the topology file. These files have been included into the CP-PAW repository and thus are available with the whole package.

The CP-PAW package is built by many different objects. The philosophy of the program structure comes from the method of object orientated programming. Objects unite variables and subroutines for special purpose. For example there is an object which contains all routines to manage the file system (paw_filehandler) or there is an object which manages the programs behavior in the case of errors (paw_error).

As mentioned above, the Universal Force Field has been included in CP-PAW earlier. All subroutines concerning classical systems are gathered in the paw_classical object. The subroutines for the input and output of data can be found in the paw_ioroutines object. These are the main objects which have to be modified to include the AMBER force field. The paw_forcefield object is a new object which contains the subroutines for handling the data needed by the AMBER force field. While writing paw_forcefield, a major intention has been to keep it flexible so that other force fields can be easily included.

In the following all program variables are written in capital letters. The data in the structure and the control files are given in data blocks. Data blocks start with an exclamation mark followed by the name of the block. Each block ends with !END. Within the blocks, new blocks can be opened or data is set in a data field. The data field is described by the name of the field, followed by an equals sign and the data itself. In the text the data field is separated with a colon from the block structure. For example the AMBER force field shall be used in a calculation. This is specified in the structure file and part of the !FORCEFIELD block which belongs to the !QMMM block. The variable is FF. In the text I will write this as !STRUCTURE!QMMM!FORCEFIELD:FF='AMBER' while in the structure file the variable will be set by:

```
1 !STRUCTURE
2     !QMMM
3         !FORCEFIELD
4             FF='AMBER'
5         !END
6     !END
7 !END
```

5.1.2. Classical Simulations with AMBER - an Overview

First I will describe the general process of a classical simulation with the AMBER force field. In the following sections I will go deeper into detail.

All data needed for the AMBER force field are provided by two input files: the parameter file and the topology file. The parameter file contains all atom types the force field can handle. Because AMBER was built for organic systems, especially proteins and DNA, there are only few atoms parameterized which

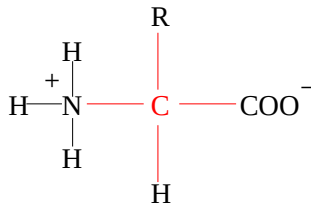


Figure 5.1: A zwitterionic amino acid. At physilocally pH-values the amino group is protonated and the carboxyl group is deprotonated. The R-group determines which amino acid it is.

are the main elements in organic systems. The parameterized elements are hydrogen, carbon, nitrogen, oxygen, sulfur and phosphorus. Each element has different force field types depending on the role the atom plays in the molecule. For example each amino acid contains at least two carbon atoms. The basic structure of an amino acid is $-NH-CHR-CO-$ which is shown in Figure 5.1. R denotes the side chain of the amino acid which characterizes it. The central carbon atom has the force field type CA the second amino acid is of type C. The first carbon atom in the side chain would be called CB and so on. Appendix C gives a small overview about amino acids.

For special combinations of atoms there are parameters for the bonds, angles and torsions between them as well as for the van-der-Waals interactions. The topology file contains informations about the structure of amino acids and deoxyriboses. Some special topologies, e.g. water, are also included. The topologies contain the information of the total charge, the bonds belonging to the residue, angles and torsions.

The user has to make sure that the system he wants to simulate is only described by residues that are part of the topology file. If he wants to simulate something else the user has to complete the topology file.

In each iteration of the classical simulation the same steps will be taken. First the neighborlist is constructed[1]. This is necessary to decide which atoms are taken into account for the non-bonded interactions and which atoms are taken into account for the bonded interactions. Then the total energy will be calculated. The total energy is the sum of all energy terms as described in equation (3.2.2). This yields the forces on the atoms. In the next step the atoms will be propagated according to these forces (and friction, if chosen). Finally the new atomic positions will be switched and the next iteration can begin. Figure 5.2 depicts the steps of each iteration cycle.

The following section provides a deeper look into the CP-PAW code to understand how the AMBER force field is implemented.

What the user has to do to use AMBER

Starting a QM/MM calculation the user has to provide the input and parameter files. Additionally, the variable !STRUCTURE!QMMM:FORCEFIELD='AMBER'

[1] To improve the performance the neighborlist sometimes is not evaluated in each iteration step. This can be easily changed in the code.

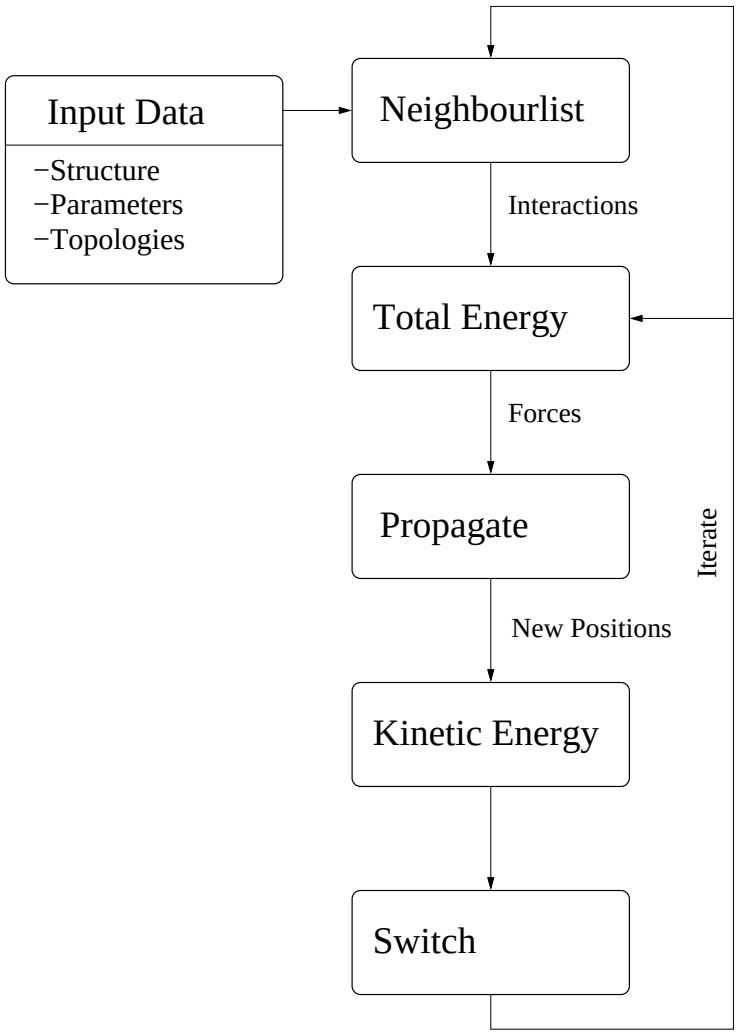


Figure 5.2: Iteration cycle of an molecular dynamics simulation.

must be set in the structure file to choose AMBER as MM force field. The program will look for the parameter and topology file. If these files are not in the standard directory[2] or if they are not specified by the variables !STRUCTURE!QMMM:PARMFILE and !STRUCTURE!QMMM:TOPFILE the program will end up in an error message. The PDB file which contains the structure data of the MM part is expected to exist in the project directory and to have the same root name with '.pdb' as suffix. Otherwise location and name of the PDB file can be specified in the !QMMM block.

5.1.3. Reading the data in paw_ioroutines

The paw_ioroutines object is responsible to read all the data needed for an QM/MM calculation with the AMBER force field for the MM part. It transfers the data to the other modules. All these things are done in the subroutine STRCIN_SOLVENT. After the data is read from the structure file the force field object reads in the parameter and the topology file.

The Level of Detail

Whenever the subroutine CLASSICAL$REPORT is called a set of informations is written to the protocol file. These informations contain the atoms of the MM system, the used potentials, the bonds, angles and torsions in the MM system. Especially for large MM systems the data becomes very large and unclear. While this data is often useful for debug information it can be neglected in most calculations. I have introduced the variable LOD (level of detail) which can switch the amount of information the report contains. Level one only gives the number of atoms, potentials, bonds and so on. Level two will additionally give the atoms' positions. Level three adds the potentials and from level four on all information is given. The variable is set in the control file by !CONTROL!QMMM:LOD=1..10. At the moment ten is the maximum value.

Data from the parameter file

First the program reads the number of bond, angle, torsion, improper torsion and van-der-Waals entries of the parameter file. After that the arrays for the corresponding data are allocated. In a second loop the file is read again and the data of the different entries will be transfered to the force field module. In this subroutine the value of the variable FORCEFIELD is checked. If it is not 'AMBER' the subroutine will stop the program with an error message that the force field is not recognized[3]. At this point it is possible to add an 'ELSE' block to read in a different type of parameter file for an other force field.

[2] The standard files are specified in the force field module. You can change them to your systems belongings if you like to.

[3] It has to be remarked that this part of the program is only used for AMBER calculations. The Universal Force Field (UFF) is treated in the paw_classical object and not in the paw_forcefield object.

Data from the topology file

Reading the topologies is more complicated than reading the parameters. Twenty essential amino acids plus special cases, some nucleotides and the patch residues have to be read in. Each of them contains a different number of atoms, bonds, angles and so on. Also the number of the residues may change if the user adds additional structures. For each of the residues there is one entry in the TOP_RES array[4]. This array is of type RES_TYPE which is defined in the force field module. RES_TYPE contains a character string for the residue name (RESNAME), and a real variable for the total charge of the residue (TOTALCHARGE). It also contains arrays for the atoms (ATOM), bonds (BOND), improper dihedrals (IMPROPER) and internal coordinates (IC)[5] of the residue. While each residue has a different number of these data entries the arrays are declared with the parameters DIMENSION(:),POINTER. After allocating the TOP_RES array with the number of residue entries in the topology file these arrays have to be nullified. In a second loop reading the topology file the atoms, bonds, torsions and ICs of each residue are count and the arrays for each part of TOP_RES are allocated. Because there is normally more than one bond pair per line in the topology file the program only counts the number of lines with the BOND keyword. The bonds will be read into an auxiliary array called BONDS. Later the data from BONDS will be copied to the TOP_RES%BOND array.

After all arrays have been allocated they will be filled with data in a third loop. The data for the atom entries are different for amino acids and for nucleotides. For that reason the variables TRESI and TRESIDUE have been introduced to control the input format. First the program reads the atom name (NAME), the force field type (ATOM) and the charge (CHARGE) of each atom. In the next step the program reads the bond entries. As mentioned above, each line will be stored in a dummy variable. Here the program only checks whether there is an even number of bond entries per line. The improper dihedrals will be read in the next step. All four atoms (ATOM1,... ATOM4) are stored in the array. The IC entries are read next. The format of the ICs is quite easy to understand. The first four entries are the atoms (ATOM1...ATOM4) taking part followed by five numbers. The first number is the distance between atom number one and two. The second number is the angle between the atoms one, two and three. The torsion between the atoms is the third number. The last two numbers are the angle between the atoms two, three and four and the distance between atoms three and four. Each residue is finished when the line begins with 'PATCH' because the program does not use the patch entries from the topology file to connect different amino acids. At the end of the subroutine the bond data from the dummy array is copied to the TOP_RES variable, the local arrays are deallocated and the subroutine FORCEFIELD_READPATCHRESIDUE is called which reads the patch entries in the topology file and adds the data to the

[4] SIZE(TOP_RES) is the sum of the number of residue entries and the number of patch residues.

[5] the IC entries obtain additional structural information in the case the amino acid is connected to another one. At the moment these informations are not used by the program.

TOP_RES array. The entries in the topology file contain only the difference between the 'normal' amino acids or deoxyriboses and the same at the end of a chain. It is also important at which end of the chain the residue stands. The name of the patch residue contains this information (N or C for amino acids and 5 or 3 for the deoxyriboses.).

The subroutine works as follows: the program checks to which original residue the patch residue belongs and makes a copy of the original entry in the TOP_RES array. Then it looks for the differences and deletes the dispensable atoms, changes the charges as well as the values for the atoms, bonds and ICs as needed. To change the data the FORCEFIELD$ADDXXXX subroutines are used. As an example the subroutine to add an atom is described in detail. The other subroutines work in the same way.

FORCEFIELD$ADDATOM In the first step the program allocates an array called OLDATOMS with the size of the original residue's atom array. Then all atoms will be written into the new array. The old atom array will then be deallocated and allocated again with the size increased by one. The data from OLDATOM is written into the new array and finally the variable NEWATOM which is an input variable of the subroutine of type TOP_ATOM_TYPE is written into the new array. The OLDATOM array is deallocated in the last step. All other ADD-subroutines work in the same way.

'END' marks the end of the topology file. The subroutine will stop here and return.

Data from the PDB file

If the user wants to make an AMBER calculation he has to provide the starting structure by a PDB file. The program uses the PDB ASCII file format as it is defined on the PDB home page (pdb.org).

In the opening of this chapter the extention of the PDB files for the CP-PAW program has been mentioned. The idea behind this is that it will be possible to use CryMolCAD (CMC) to build the input PDB files. CMC is a 3D viewer and builder for atomic structures. It is developed by the IT team in our institute. It is planned to set the flags (Q labels the QM part, L labels the link atoms) in the PDB file using the graphical interface. Therefore, it can be easily decided which atoms belong to the QM system, which atoms are link atoms and so on. All input files needed to do a QM/MM calculation will be written by CMC. But as long as this part of CMC is not finished, the user has to do these things by hand.

The user can also freeze parts of the system by simply setting the 'F' flag. This fixes the atoms in space.

PDB files usually are without hydrogen atoms. Word et. al. presented the program 'reduce' [88, 89] which is able to add the missing hydrogen atoms to the PDB-file. The program also optimizes the positions of the side groups in the amino acids by rotating them, so that the hydrogen atoms can build reasonable hydrogen bonds with the environment. The PDB file generated from 'reduce' can be used as input file in AMBER_MD and in CP-PAW.

In QM/MM calculations the structure of the QM part has to be set in the STRC input file, too. To get the coordinates of the participating atoms into the CP-PAW input format the tool paw_tostrc can be used. I expanded the tool so it is able to read the bare PDB files and add the missing hydrogen atoms or it reads a PDB file which has been processed by 'reduce'. In this case the input file must have the suffix '.pdb_red'.

The data of the final PDB file for a QM/MM calculation will be read by the subroutine FORCEFIELD$READ_MMSTRC. PDB files often contain more data than the CP-PAW program can use, e.g. the data about the secondary, tertiary and quartary structure of the protein. It only needs the primary structure. Therefore the subroutine only takes care about lines with 'ATOM', 'HETATM' or 'CONECT' at the beginning. An example for a PDB file is given in Section 5.3.

All data from the ATOM and HETATM lines is read into the MMATOM array which is of PDB_ATOM_TYPE type. This type contains the following data:

- KEYWORD is a character(6) variable which contains the first entry, ATOM or HETATM.
- ID is the serial number of the atom. Important: ID is not necessarily the number of the atom in the array. For example this is important if there are CONECT entries in the structure. They use the ID number and when the program looks for the atoms to form the bonds, they have to be translated because the bonds will be formed between the array IDs. When writing the PDB_OUT file the program takes care about these numbers and translates them back into ID numbers again.
- NAME (character(5)) is the name of the atom. It should not be used to get the element name. This can be easily seen with the following example: each center carbon atom of an amino acid has the name 'CA' which is easily mixed up with the element calcium.
- RESNAME (character(4)) is the name of the residue.
- CHAINID is a character(1) variable which identifies the chain, the amino acid belongs to. Amino acids with the same CHAINID will be connected between the final carbon atom of the actual amino acid and the first nitrogen atom of the next amino acid. All none amino acid residues should be of HETATM type and have the same CHAINID (often I use X as CHAINID).
- RESSEQ contains the residue sequence number.
- into R the program writes the x, y and z coordinates of the atom. The coordinates in the PDB file are in Ångstrom by default. The program converts them into Bohr radii at once.
- OCCUPANCY, TEMPFACTOR and SEGID is not used by the CP-PAW code in the moment.

- ELEMENT is a character(2) variable for the element type of the atom. Not all PDB files use this field, so the program has to make sure that this variable is filled in the right way. The QMMM object needs this variable to determine the vdW-radius of the link atoms. If the element data field in the PDB file is empty the program gets the element name as

```
1   MMATOM(I)%ELEMENT = ADJUSTR(MMATOM(I)%NAME(1:2))
```

  In most cases this will work fine because more than 95% of the atoms in a PDB file are hydrogen, carbon, nitrogen or oxygen. If there are any other elements they should be labeled in the PDB file[6].

- FLAG is a character(1) variable which has been introduced in the PDB format used by the program. At the moment the flag can be set to 'F', 'Q' or 'L'. Using the 'F' flag tells the program that the atom is fixed and will not be propagated by the classical object. 'Q' labels the atom as part of the QM part. 'L' labels a link atom.

- QMNAME is a character(12) variable which contains the name of the atom in the structure file of the CP-PAW QMMM calculation. The user is free to choose the name but it is recommended to use it in the following way: the first two characters are the element name (elements with only one character use an underscore as second character). Then the missing characters of the atoms name are added. After the next underscore the ID number of the atom follows. For example the center carbon atom of an amino acid will get the name 'C_A_23'. If there is a link atom then QMNAME contains the QJOINT atom.

- LNAME is a character(7) variable which contains the name of the QATOM belonging to the link bond. The name has to be the same as it is set in the structure file.

In the next step the connect informations will be read. If there are only residues which are defined in the topology file then there is no need to take care about any bonds in the system. But for all other atoms the bonds must be defined (e.g. the bonds between the P-cluster of Nitrogenase and the enzyme are described in this part of the PDB file.). As mentioned above the program has to reassign the numbers of the bonded atoms because the program uses the array index and not the ID variable of the CONECT entries. The form of a connect entry is the keyword (CONECT) followed by five integer numbers with a maximum of 5 digits for each (FORTRAN format string is '(A5,5I5)'). The first number is the central atom the other four are possible bond partners. Not all numbers have to be filled with data.

In the last part of the subroutine the program treats some special cases. During the work on this thesis the program 'reduce' written by Word et. al. [88] was mainly used to complete the PDB files. Standard PDB files do not have

[6] At the end of the STRCIN_SOLVENT subroutine are some lines in the code which are out-commentated. Remove the exclamation marks, compile the code and start it again to make sure that the data transferred to the classical object is ok.

any informations about hydrogen atoms. The paw_tostrc tool is able to add them and define the bonds. A better result is yielded by 'reduce'. But 'reduce' leads to some problems with wrong names. Some problems have been identified yet and they are described below. One of these problems is that these added hydrogen atoms have names which do not fit into the space the PDB format has reserved for names. For example, using valine one gets a hydrogen atom 'HG13'. The last digit will not fit into the PDB file. The trick here is that 'reduce' uses the empty column before the name columns, which results in an atom name like '3HG1'. The program checks whether the empty column is filled with a number and adds this number at the end of the atom name. Another problem can be that the element column is empty. In this case the element variable is filled with the first two characters from the MMATOM%NAME variable. A last workaround is that some PDB files with only one protein chain do not use the CHAINID variable. In this case the variable is filled with 'A' by default.

Now all the data the user provided for a calculation is read. The next step is to transfer the data in the right form to the classical object.

5.1.4. Transferring the data to the classical object

After the data is read from the input files paw_ioroutines, processes the data to the classical object. This is done in three steps: first the atoms will be processed then the links and finally the bonds. For each step an IF statement decides whether the data will be processed for an UFF or for an AMBER calculation. At the moment the main difference is that the data for UFF calculations is completely found in the structure file while the data for the AMBER calculations comes from an PDB file.

Atoms

The complete data from the PDB file is stored in the variable MMATOM of the force field module. The program transfers the data to the local variables MATOM (for the MM part) and SATOM (for the shadow). The QM system is read from the structure file and stored in the atom list.

First, the names of the atoms have to be constructed. This is done by:

```
1  MATOM(IATM)%NAME = TRIM(ADJUSTL(MMATOM(IATM)%NAME(1:LEN_TRIM(&
2          &          MMATOM(IATM)%NAME))))//'_'//TRIM(ADJUSTL( &
3          &           ITOS(MMATOM(IATM)%ID)))
```

The name of each atom in the MATOM array has to be unique. Thus, it is constructed by the original name of the atom in the PDB file plus an underscore and the atoms ID number[7]. The function ITOS translates an integer variable into a string. It is defined in the paw_strings object.

In the next step the program looks for atoms with a 'Q' flag. These atoms belong to the QM part. For such an atom the variable IATQ is stored and the variable IATS is increased by one so the data (name and element) can be written into SATOM(IATS). After that the QMSATOM variable is set for all three subsystems. Each atom of each subsystem (QATOM, MATOM, SATOM)

[7] Remember: ID number is not the number of the atom in the array!

has the QMSATOM variable which is an integer variable. It points to the same atom in another subsystem. The MM atoms point to the atoms in the shadow, the shadow atoms point to the atoms in the QM subsystem and the QM atoms point to the atoms in the MM subsystem. If the number of one atom in any subsystem is known, access to the same atom in all other subsystems is possible. If the atom is only found in the MM part (i.e. it belongs to the environment) the QMSATOM variable is set to zero.

The force field type and the charge come from the force field files. To get access to them the program looks for the residue name. The subroutine FORCEFIELD$GETRESNUMBER returns an integer number which points to the residue in the TOP_RES variable. From this the program gets the force field type of the atom and the charge. Note, that the charges in the program are in electron charges (i.e. one electron carries the charge 1). Therefore, they are multiplied with minus one. Whenever the program writes out charges for the protocol file it transforms them back. To avoid misunderstandings it always writes 'Q[E]' in this case. The coordinates of the atoms are also transformed to atomic units.

In the last steps the program sets the masses of the atoms and checks the flag of the atom. If the flag is 'F' the atom is freezed.

Links

All link bonds are marked by the L-flag in the PDB file. Figure 5.3 explains the name policy of the atoms which take part in a link bond. The program has

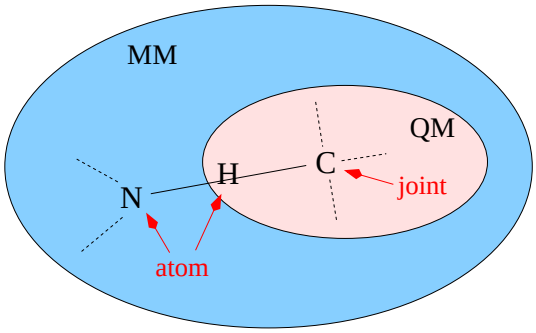


Figure 5.3: Name policy of the link atoms.

to find out which atoms belong to a link bond and which role they play. The data will be written into the LINK variable which is of type LINK_TYPE. The LINK_TYPE is defined as

```
1       TYPE LINK_TYPE
2         INTEGER(4)          :: MJOINT
3         INTEGER(4)          :: QJOINT
4         INTEGER(4)          :: SJOINT
5         INTEGER(4)          :: MATOM
6         INTEGER(4)          :: QATOM
7         INTEGER(4)          :: SATOM
8       END TYPE LINK_TYPE
```

and contains the number of the parting atoms in all subsystems. At the end of this part the shadow dummy atom is defined. For the dummy atom a hydrogen atom is used (atom name 'H') which has 'L1' as force field type. The force field type 'L1' is defined in the parameter file. The user has to make sure that there exist the right bonds, angles and vdW-parameters in the parameter file as needed. For links as described in this work, the parameters are already stored in the parameter file.

Bonds

Using the UFF force field in the QM/MM scheme, all bond informations have to be given in the structure input file. For AMBER as choice for the MM part, the topology file contains the bond information and the user does not have to take care about it. The idea behind AMBER is that the starting structure is constructed from a defined sequence which gets the missing information from the topology file. Therefore, each residue contains bonds with '+N' or '-C' atoms. These atoms are the connecting atoms of the neighbor residues. This does not fit into the concept of the implementation in the CP-PAW code so bond entries with such atoms are ignored.

To get the right number of bonds, the bonds are counted from the information in the topology file. Additionally the bonds between amino acids and the bonds defined in the CONECT entries have to be added.

The program starts with a loop over all atoms of the MMATOM variable. First, the program determines the residue to which the first atom belongs to. This is done by:

```
1     RESNAME= MMATOM(IAT)%RESNAME
2     CALL FORCEFIELD__GETRESNUMBER(RESNAME,IVAR)
3     IBONDM= SIZE(TOP_RES(IVAR)%BOND)
```

The variable IBONDM carries the information of the number of bonds in the residue the first atom belongs to. If there are any bonds with atoms '+N' or '-C', they will be subtracted. In the next step the keyword variable of the atom is checked. If it is 'ATOM' then the atom belongs to an amino acid and again one bond is added for building the connection to the next amino acid. If the keyword is 'HETATM' then only IBONDM will be added to the number of total bonds (which is stored in the variable 'NBONDM'). The last step in the loop is to increase the loop variable IAT by the number of atoms in the residue:

```
1     IVAR = SIZE(TOP_RES(IVAR)%ATOM)
2     IAT = IAT + IVAR
```

This is a jump to the first atom of the next residue. The program can now check if the next residue belongs to a different chain than the original residue or if the original residue is the last residue. In both cases one bond is subtracted.

Finally the program takes care of the CONECT entries and counts the bonds that are specified there. The array MBOND will be allocated with the number of total bonds. In a final loop over all atoms the MBOND variable will be filled with numbers from atoms connected by a bond. Again the program looks for the residue of the first atom and gets the bonding information. If both atoms belonging to a bond are found, their numbers are written into the MBOND

variable. The bond order will be set to one[8]. In the next step the amino acids will be connected together and finally the CONECT bonds are set.

Now all the data from the input file is processed and written into local variables. These variables are the same for UFF and for AMBER calculations. In a last step the shadow bonds are resolved. This is the same code for each force field. The subroutine STRCIN_SOLVENT_SETM is called twice to transfer the data to the classical object (once for the MM part and once for the shadow.).

5.1.5. Using the AMBER parameters in the classical object

At this point all the data from the input files has been read and transfered to the classical object. Also the force field type which shall be used is set. The subroutine CLASSICAL$INITIALIZE calls the subroutine CLASSICAL_FORCEFIELDSETUP which constructs the potentials. When creating a potential the program looks which force field is selected and calls the appropriate subroutine to create the potential.

Details of the force field

The program uses the Cornell et. al. force field for MM calculations as it is described in [21]. The total energy is calculated as described in section 3.2 by equation 3.2.2. The next step is to take a closer look how the potentials are constructed in the program. This starts with the interactions between bonded atoms.

The bond parameters are provided by the subroutine FORCEFIELD$AMBER_BONDPARMS. It shall be remarked that UFF uses the value $\frac{K}{2}$ for the force constant while AMBER uses K. In order to use the same subroutines to evaluate the potential, K is multiplied by a factor two so that the subroutine CLASSICAL_BONDPOTA can be used for both force fields. The potential is created by

```
1       DO I=1,POT%NX
2         X=POT%X1+POT%DX*DBLE(I-1)
3         R=1/X
4         DRDX=-R**2
5         POT%VAL(I)=0.5D0*KIJ*(R-RIJ)**2
6         POT%DER(I)=KIJ*(R-RIJ)*DRDX
7       ENDDO
```

The same procedure is done for the angles. First the program gets the parameters of an angle by calling FORCEFIELD$AMBER_ANGLEPARMS. The subroutine FORCEFIELD$AMBER_ANGLEPOTA constructs the potential out of these parameters. This is done by

```
1       DO I=1,POT%NX
2          X=POT%X1+POT%DX*DBLE(I-1)
3          THETA=DACOS(X)
4          POT%VAL(I)=KIJK*(THETA-THETA0)**2
5          POT%DER(I)=2*KIJK*(THETA-THETA0)*(-1.d0/sin(THETA))
6       END DO
```

[8] The bond order is not used by AMBER so I set it always to one.

As discussed in section 3.1.2, it is reasonable to use a Fourier expansion rather than a harmonic expression. Because force field parameters are only valid together with the whole force field, the harmonic expression will be used by default in the CP-PAW program. But the user can choose which angular potential he wants to use. If the user wants to use a Fourier expression for the potential the force field parameter will be multiplied by two. n is set to one. To switch between the two potentials the variable TFOURIER variable can be set in the code. It is located in the CLASSICAL_FORCEFIELDSETUP subroutine. By default the variable is set to false. In this case the harmonic angular potential will be used. A first test with the tri-peptide system (the molecule is described in section 5.3) shows two different results for both expressions. The energy difference is 0.5kJmol^{-1}. Figure 5.4 shows the resulting structures. The blue system was calculated with the Fourier expression, the red one with the harmonic expression.

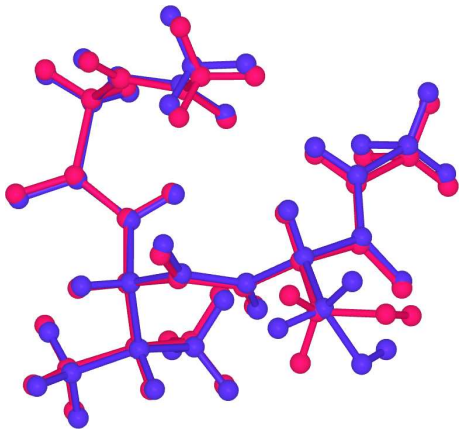


Figure 5.4: Two different final structures of the tri-peptide test molecule. The blue system was calculated with a Fourier expression for the angular potential, the red one was calculated with a harmonic expression.

The parameters for the torsion potential are provided by the FORCEFIELD$-AMBER_TORSIONPARMS subroutine. The parameters are the barrier V_n, the periodicy n and a phase offset γ. The last parameter changes the sign of the cosine. In the UFF formulation of the torsion potential the expression looks like

$$E_\phi = \frac{1}{2} V_\phi \left[1 - \cos n\phi_0 \cos n\phi\right] \tag{5.1.1}$$

In the AMBER related subroutine the potential is constructed by:

$$E_n = \frac{1}{2} V_n \left[1 - \cos\left(n(\theta - \theta_0)\right)\right] \tag{5.1.2}$$

This is possible because the parameter γ (which is the same as θ_0) only has the values zero or π. The potential is created by:

```
1        DO I=1,POT%NX
2           X=POT%X1+DBLE(I-1)*POT%DX
```

```
3           X=MIN(1.D0,X)
4           PHI=ACOS(X)
5           POT%VAL(I)=0.5D0*V*(1.D0-COS(DBLE(NJK)*(PHI-PHI0)))
6           IF(I.NE.1.AND.I.NE.POT%NX) THEN
7             POT%DER(I)=-0.5D0*V*DBLE(NJK)*SIN(DBLE(NJK)* &
8           & (PHI-PHI0))/SIN(PHI)
9           ELSE    ! AVOID DIVIDE BY ZERO AT THE END POINTS
10            POT%DER(I)=-0.5D0*V*DBLE(NJK)**2*COS(DBLE(NJK)*&
11          & (PHI-PHI0))/COS(PHI)
12          END IF
13        ENDDO
```

At this point all interactions between bonded atoms have been discussed. What is left are the non-bond interactions, which are the coulomb interaction and the van-der-Waals interaction. In the next step the coulomb potential is evaluated. The coulomb potential in AMBER is described by a simple $1/r$ potential while UFF uses a coulomb potential derived by the error function. The subroutines CLASSICAL_COULOMBPOTA and CLASSICAL_COULOMBPOTB are used to create these potentials. The potential in CLASSICAL_COULOMBPOTA is created by

```
1         DO I=1,POT%NX
2           X=POT%X1+POT%DX*DBLE(I-1)
3           IF(X.NE.0.D0) THEN
4             POT%VAL(I)=X
5             POT%DER(I)=1.d0
6           ELSE
7             POT%VAL(I)=0.D0
8             POT%DER(I)=1.D0
9           END IF
10        ENDDO
```

while the potential in CLASSICAL_COULOMBPOTB is created by

```
1         DO I=1,POT%NX
2           X=POT%X1+POT%DX*DBLE(I-1)
3           IF(X.NE.0.D0) THEN
4             CALL LIB__ERFR8(1.D0/X,ERFX)
5             POT%VAL(I)=X*ERFX
6             POT%DER(I)=ERFX-2.D0/(X*SQRT(PI))*EXP(-1.D0/X**2)
7           ELSE
8             POT%VAL(I)=0.D0
9             POT%DER(I)=1.D0
10          END IF
11        ENDDO
```

Originally, the exclusions[9] have been ignored when calculating the coulomb interaction in the UFF scheme. Therefore, a potential has been introduced, which is generated by the error function. It gives an accurate behavior of the potentials at short distances. Figure 5.5 shows the value and the derivative of the used coulomb potentials. However, using the same potential with the AMBER force field parameters leads to an non-physical behavior in some cases. One important example is the calculation of a water dimer which always ended in a planar structure. Applying the $1/r$ potential in the AMBER calculations solved this problem.

The last interaction is the van-der-Waals interaction. Here special care has to be taken about water molecules. The TIP3P water model [72] is part of

[9] The exclusions contain the atoms which do not have any non-bonded interactions. They will be explained later in more detail.

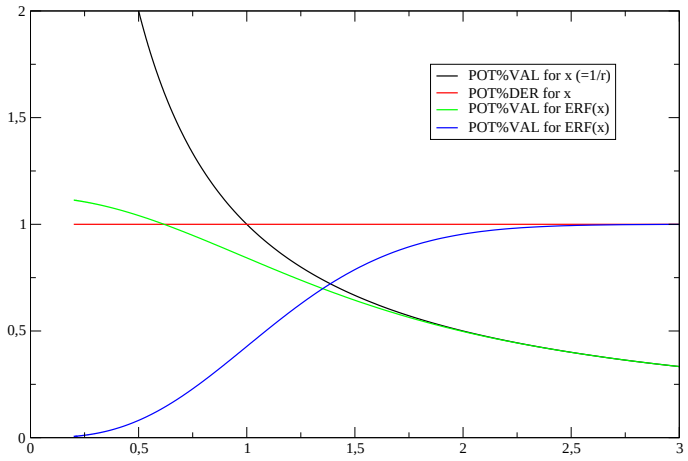


Figure 5.5: Value and derivative of the Coulomb potentials which is used in the AMBER implementation in the CP-PAW package. For larger distances both potentials approach each other. For smaller distances they diverge.

the AMBER force field. In this model the water molecule is described as a rigid triangle. The distance between the atoms and the angles remains constant during the simulation. Also the van-der-Waals interaction is considered only between the oxygen atoms. In the FORCEFIELD$AMBER_NONBONDPARMS subroutine this is managed by:

```
1  TCHK=.TRUE.
2  IF( (TRIM(ADJUSTL(ATOM1)).EQ.'OW'.AND.TRIM(ADJUSTL(ATOM2)).EQ.'HW').
       OR. &
3    & (TRIM(ADJUSTL(ATOM1)).EQ.'HW'.AND.TRIM(ADJUSTL(ATOM2)).EQ.'OW').
         OR. &
4    & (TRIM(ADJUSTL(ATOM1)).EQ.'HW'.AND.TRIM(ADJUSTL(ATOM2)).EQ.'HW')
           &
5    & ) THEN
6     TCHK=.FALSE.
7     RIJ=0.d0
8     DIJ=0.d0
9    RETURN
10 END IF
```

When the subroutine returns TCHK as false the variable NONBOND(IAT1,IAT2) becomes minus one. In this case the program does not ask for the value of the potential. Finally, the potential is created by the following code:

```
1        SVAR1= DIJ * RIJ**12
2        SVAR2= 2* DIJ * RIJ**6
3        DO I=1,POT%NX
4            IF(I.NE.1) THEN
5               X=POT%X1+POT%DX*DBLE(I-1)
6               POT%VAL(I)=SVAR1*x**12 - SVAR2*x**6   ! x=1/r
7               POT%DER(I)=12.d0*SVAR1*x**11 - 6.d0*SVAR2*x**5
8            ELSE
9               POT%VAL(I)=0.D0
10              POT%DER(I)=0.D0
11           END IF
```

```
12        ENDDO
13        POT%VAL(NX)=POT%VAL(NX-1)+0.5D0*POT%DER(NX-1)*POT%DX
14        POT%DER(NX)=0.D0
```

At this stage all potentials needed to calculate the system are created and the evaluation of the total energy can begin. In each iteration the subroutine CLASSICAL$ETOT is called. This subroutine calculates each energy term in equation (3.2.2) and sums them up. The appropriate subroutines to get the energies are the same which are used by the UFF implementation. The only change is made in the subroutine CLASSICAL_ECOULOMB. This subroutine is used to get the value of the non-bonded interactions of two atoms. In the first step, the atoms participating in non-bonded interactions have to be defined. This is done by the subroutine CLASSICAL_EXCLUSIONS which scans the neighborlist in order to find atoms which are bonded directly (1-2 interaction) or which are bonded by a common atom (1-3 interaction). These atoms are excluded from non-bonded interactions in both force fields, AMBER and UFF. The difference between the two schemes is that in the AMBER force field the 1-4 interaction is treated in a special way. As already mentioned in section 3.1.2 the coulomb interaction of 1-4 bonded atoms is scaled by a factor 1/1.2 and the van-der-Waals interaction of these pairs is scaled by 1/2.0. Therefore, it is possible to use the same subroutine for UFF and AMBER calculations. Only small modifications are necessary. An additional part checks the possible 1-4 interactions of the atoms. The variable TONEFOUR has been introduced which is set TRUE if the atoms are the edges of a torsion. This is done by:

```
1       TONEFOUR=.FALSE.
2       DO I=1,SIZE(MD%TORSION)
3          ITOR1=MD%TORSION(I)%IAT1
4          ITOR2=MD%TORSION(I)%IAT4
5          IF( (ITOR1.EQ.IAT1.AND.ITOR2.EQ.IAT2).OR.&
6                &  (ITOR1.EQ.IAT2.AND.ITOR2.EQ.IAT1) ) THEN
7             TONEFOUR=.TRUE.
8             EXIT
9          END IF
10      END DO
```

The next modification has become necessary after the introduction of the $1/r$ potential. When the coulomb interaction is described by that potential, the QM/MM calculations do not converge. The QM region is strongly polarized and the wave functions cannot be orthogonalized; the calculation stops with an error message.

The energy of a coulomb interaction between two atoms is calculated by[10]:

```
1       IF(.NOT.TEXCL.AND.MD%FF.EQ.'AMBER') THEN
2          RC=1.D0
3          CALL VALUE(POT(1),RC*X,G,DGDX)
4          IF(TONEFOUR.AND.MD%FF.EQ.'AMBER') THEN
5             G=G * (1.d0/1.2d0)
6             DGDX= DGDX * (1.d0/1.2d0)
7          END IF
8          G=G/RC
9          E=E+Q1*Q2*G
```

[10] This will soon be changed in the code. The variable TONEFOUR and SCALEONEFOUR will be set when the potentials are created. CLASSICAL$ECOULOMB only has to decide which potential should be used. Look at the paw_classical object for details.

```
10          FAC=FAC+Q1*Q2*DGDX
11          V(IAT1)=V(IAT1)+Q2*G
12          V(IAT2)=V(IAT2)+Q1*G
13        ELSE
14          RC=4.d0
15          CALL VALUE(POT(2),RC*X,G,DGDX)
16          IF(TONEFOUR.AND.MD%FF.EQ.'AMBER') THEN
17            G=G * (1.d0/1.2d0)
18            DGDX= DGDX * (1.d0/1.2d0)
19          END IF
20          G=G/RC
21          E=E+Q1*Q2*G
22          FAC=FAC+Q1*Q2*DGDX
23          V(IAT1)=V(IAT1)+Q2*G
24          V(IAT2)=V(IAT2)+Q1*G
25        END IF
```

It can be easily seen that the simple $1/r$ potential is only used for atom pairs in an AMBER calculation which are not excluded. In all other cases (UFF calculation or excluded atoms in an AMBER calculation) the coulomb potential derived from the error function is used. For both potentials the 1-4 interaction is scaled when the AMBER force field is chosen.

The subroutines which calculate the other energy terms have not been changed. UFF and AMBER can use the same routines to evaluate the energy terms.

5.2. AMBER_MD - a stand alone MD program

AMBER_MD is a stand alone molecular dynamics simulation program which uses all above described routines from the CP-PAW program package. In the latest version it uses the parameter and topology files from the Cornell et. al. force field. PDB files serve as structure input files. The structures can be optimized by friction molecular dynamics. Also a thermostat is implemented to perform simulations at finite temperature. The structure data is written to an xyz-file in each time step. However, the user can define how many steps shall be skipped in order to keep the size of movie file small. The file can be used to analyze vibrational data. Figure 5.6 shows an example control file. In the !GENERIC block (lines 2-4) the variable START defines whether a simulation starts from the beginning or an old calculation shall be continued. Using START=F a file `case.pdb_out` must exist in the directory. TRACE enables debug informations and shall not be used in calculations. LUNIT defines the conversion factor for the length units. By default input files are in Ångstrom. In this case the factor presented in the figure has to be used.

The !RDYN block (lines 6-10) defines the dynamics of the system, i.e. the applied friction (FRIC), the number of iterations (NSTEP) and the time step (DT). A time step of 10.335533 a.u. stands for 0.25 fs.

In the !THERMOSTAT block (lines 12-15) a thermostat can be used to simulate systems at a finite temperature[11]. The target temperature (T[K]) and the period of the thermostat (FREQ[THZ]) can be set. The !MOVIE block (lines 17-19) allows the user to skip some iterations while writing the movie file.

[11] Note, that using a thermostat the friction has to be zero in the !RDYN block.

```
!AMBER
  !GENERIC
     START=F TRACE=F LUNIT=1.8897259926
  !END

  !RDYN
     FRIC=0.0008
     NSTEP=40000
     DT=10.335533
  !END

  !THERMOSTAT
    T[K]=300.
    FREQ[THZ]=5.
  !END

  !MOVIE
    SKIP=10
  !END

  !FORCEFIELD
      LOD=1
      MMFILE='case.pdb'
      PARMFILE='/home/shemmen/PAW/forcefields/amber/cornell_all.prm'
      TOPFILE='/home/shemmen/PAW/forcefields/amber/cornell_all.rtf'
      !LATTICE
         T=100.   0. 0. 0. 100. 0. 0. 0. 100.
      !END
  !END
!END
!EOB
```

Figure 5.6: Example of an input control file for AMBER_MD.

Finally, the !FORCEFIELD block (lines 21-29) defines the forcefield input files (MMFILE for the starting structure, PARMFILE is the parameter file and TOPFILE is the topology file). The variable LOD defines the level of detail for the output. It has already been described in Section 5.1.3. The lattice block defines the box of the MM system.

The AMBER_MD program is basically used for testing systems. During the development of the AMBER implementation into the CP-PAW program package the AMBER_MD program was used to test all new features. Any possible errors coming from the QM/MM implementation are completely ignored. The MD program can also be used to pre-optimize MM structures which will be used in QM/MM calculations. Also it appears to be very useful when testing new topologies or force field parameters.

5.3. The construction of input files

A QM/MM simulation in the CP-PAW program package using the AMBER force field requires some special preparations of the input files. In this section these preparations are described in detail. A small peptide shall be simulated.

The system will be set up as follows: The complete peptide consists of three amino acids. The sequence is GLY-VAL-SER. The valine amino acid in the center of the peptide shall be defined as QM system. Thus, the QM part is connected by two link atoms to the MM part. While stand-alone amino acids are polarized and have charged termini, both ends of the chain are capped by special groups. Therefore, a pdb-file has been prepared with five amino acids. The sequence is ARG-GLY-VAL-SER-GLN[12]. While there is a lack of hydrogen atoms in PDB-files, the atoms must be extra added. This can either be done by the tool paw_tostrc or by 'reduce'. Both tools have been discussed in section 5.1.3. Using 'reduce' the command is

```
1    reduce input.pdb > output.pdb
```

for the conversion tool of the CP-PAW program package it is:

```
1    paw_tostrc.x input.pdb
```

The tool paw_tostrc yields a CP-PAW input format. It has to be reconverted to a PDB format. Therefore, 'reduce' is the tool of choice.

In the next step the termini will be changed. The N-terminus will be acetylated. The name of this residue is ACE ($COCH_3$). The C-terminus is capped by an N-methyl-amide group ($NHCH_3$). In both cases the needless atoms have to be deleted (red ones) and the remaining atoms have to be renamed (green ones). Figure 5.7 illustrates the change from ARG to ACE and from GLN to NME. The new PDB-file can be used as input file for calculations with the AMBER_MD program. In order to use it as input structure in a QM/MM calculation there are still some things left to be done.

All atoms which belong to the QM part have to be copied into another PDB-file. The file shall be named `case.pdb_red`. Calling

```
1    paw_tostrc.x case.pdb_red
```

[12] The five amino acids have been chosen arbitrarily from the 1M1N PDB file.

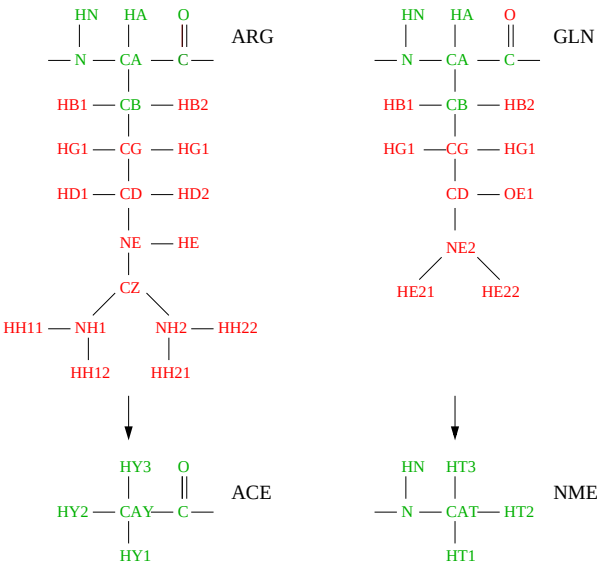


Figure 5.7: Construction of the terminal groups ACE and NME in peptide chains. The figure shows the example ARG ⟶ ACE and GLN ⟶ NME. Green atoms remain and are renamed when necessary. The red atoms are deleted.

creates a prestrc-file which can be used as source for the structure file. The structure file for the CP-PAW calculation can be set as usual.

In the next step the inner system has to be marked in the PDB file. As already mentioned the PDB file format has been extended for QM/MM calculations. The QM part can be labeled by setting the Q flag in column 81. The exact atom name as it is used in the CP-PAW structure file has to be written in columns 83 to 91 for all QM atoms. Connecting the inner with the outer system is done via link atoms. The link atoms can be placed at the position of the M1 atoms. In the structure file they are often labeled as 'H_dx' where x is an integer number. Using the labels described in Figure 5.3, the MM-atom gets the link flag. The name of the MMJOINT atom is written in column 83 to 91, the name of the QMATOM is written in column 96 to 102.

Following these instructions the PDB input file should look like:

```
ATOM      1  HY2 ACE A 187     -16.951  -9.251   0.259  1.00 24.54           H
ATOM      2  CAY ACE A 187     -18.007  -9.921  -0.474  1.00 23.72           C
ATOM      3  C   ACE A 187     -17.838 -11.432  -0.383  1.00 24.58           C
ATOM      4  O   ACE A 187     -16.733 -11.988  -0.536  1.00 23.74           O
ATOM      5  HY1 ACE A 187     -17.989  -9.538  -1.948  1.00 20.77           H
ATOM     12  HY3 ACE A 187     -18.961  -9.609  -0.023  1.00 23.72           H
ATOM     24  N   GLY A 188     -18.978 -12.099  -0.187  1.00 23.66           N
ATOM     25  CA  GLY A 188     -19.025 -13.543  -0.083  1.00 21.73           C  L C__26         H_d25
ATOM     26  C   GLY A 188     -18.366 -13.983   1.204  1.00 24.01           C  Q C__26
ATOM     27  O   GLY A 188     -18.132 -13.187   2.122  1.00 27.34           O  Q O__27
ATOM     28  HN  GLY A 188     -19.870 -11.655  -0.099  1.00 23.66           H
ATOM     29 1HA  GLY A 188     -20.069 -13.889  -0.109  1.00 21.73           H
ATOM     30 2HA  GLY A 188     -18.514 -13.999  -0.943  1.00 21.73           H
ATOM     31  N   VAL A 189     -18.040 -15.270   1.289  1.00 23.90           N  Q N__31
ATOM     32  CA  VAL A 189     -17.478 -15.815   2.522  1.00 21.98           C  Q C_A_32
```

```
ATOM     33  C   VAL A 189     -16.086 -16.389   2.356  1.00 19.85           C  Q C__33
ATOM     34  O   VAL A 189     -15.475 -16.816   3.329  1.00 19.65           O  Q O__34
ATOM     35  CB  VAL A 189     -18.412 -16.905   3.076  1.00 22.40           C  Q C_B_35
ATOM     36  CG1 VAL A 189     -19.771 -16.265   3.386  1.00 24.97           C  Q C_G1_36
ATOM     37  CG2 VAL A 189     -18.606 -18.035   2.070  1.00 25.55           C  Q C_G2_37
ATOM     38  HN  VAL A 189     -18.150 -15.932   0.548  1.00 23.90           H  Q H__38
ATOM     39  HA  VAL A 189     -17.392 -14.972   3.223  1.00 21.98           H  Q H_A_39
ATOM     40  HB  VAL A 189     -17.961 -17.332   3.984  1.00 22.40           H  Q H_B_40
ATOM     41 1HG1 VAL A 189     -20.455 -17.029   3.784  1.00 24.97           H  Q H_G11_41
ATOM     42 2HG1 VAL A 189     -19.640 -15.467   4.132  1.00 24.97           H  Q H_G12_42
ATOM     43 3HG1 VAL A 189     -20.195 -15.839   2.465  1.00 24.97           H  Q H_G13_43
ATOM     44 1HG2 VAL A 189     -19.277 -18.795   2.497  1.00 25.55           H  Q H_G21_44
ATOM     45 2HG2 VAL A 189     -19.048 -17.633   1.147  1.00 25.55           H  Q H_G22_45
ATOM     46 3HG2 VAL A 189     -17.633 -18.493   1.840  1.00 25.55           H  Q H_G23_46
ATOM     47  N   SER A 190     -15.529 -16.312   1.150  1.00 21.42           N  L C__33        H_d47
ATOM     48  CA  SER A 190     -14.261 -16.949   0.830  1.00 21.68           C
ATOM     49  C   SER A 190     -13.879 -16.484  -0.554  1.00 23.61           C
ATOM     50  O   SER A 190     -14.698 -15.836  -1.233  1.00 26.43           O
ATOM     51  CB  SER A 190     -14.426 -18.454   0.780  1.00 23.93           C
ATOM     52  OG  SER A 190     -15.394 -18.802  -0.225  1.00 17.91           O
ATOM     53  HN  SER A 190     -15.938 -15.815   0.385  1.00 21.42           H
ATOM     54  HA  SER A 190     -13.506 -16.692   1.588  1.00 21.68           H
ATOM     55 1HB  SER A 190     -14.750 -18.830   1.762  1.00 23.93           H
ATOM     56 2HB  SER A 190     -13.461 -18.932   0.555  1.00 23.93           H
ATOM     57  HG1 SER A 190     -15.499 -19.796  -0.257  1.00 17.91           H
ATOM     58  N   NME A 191     -12.736 -16.959  -1.054  1.00 20.09           N
ATOM     59  CAT NME A 191     -12.367 -16.656  -2.409  1.00 19.87           C
ATOM     60  HT2 NME A 191     -13.304 -17.276  -3.443  1.00 18.55           H
ATOM     62  HT3 NME A 191     -10.895 -17.074  -2.650  1.00 19.80           H
ATOM     67  HN  NME A 191     -12.086 -17.530  -0.552  1.00 20.09           H
ATOM     68  HT1 NME A 191     -12.465 -15.569  -2.544  1.00 19.87           H
```

The corresponding structure file in the CP-PAW input format looks like:

```
!STRUCTURE
   !LATTICE T=    0.00000   9.50000   9.50000
                  9.50000   0.00000   9.50000
                  9.50000   9.50000   0.00000
   !END

   .....

   !ATOM NAME='C__26'    R= -18.366   -13.983     1.204   !END
   !ATOM NAME='O__27'    R= -18.132   -13.187     2.122   !END
   !ATOM NAME='H_d25'    R= -19.025   -13.543    -0.083   !END

   !ATOM NAME='N__31'    R= -18.04000 -15.27000   1.28900 !END
   !ATOM NAME='C_A_32'   R= -17.47800 -15.81500   2.52200 !END
   !ATOM NAME='C__33'    R= -16.08600 -16.38900   2.35600 !END
   !ATOM NAME='O__34'    R= -15.47500 -16.81600   3.32900 !END
   !ATOM NAME='C_B_35'   R= -18.41200 -16.90500   3.07600 !END
   !ATOM NAME='C_G1_36'  R= -19.77100 -16.26500   3.38600 !END
   !ATOM NAME='C_G2_37'  R= -18.60600 -18.03500   2.07000 !END
   !ATOM NAME='H__38'    R= -18.15000 -15.93200   0.54800 !END
   !ATOM NAME='H_A_39'   R= -17.39200 -14.97200   3.22300 !END
   !ATOM NAME='H_B_40'   R= -17.96100 -17.33200   3.98400 !END
```

```
   !ATOM NAME='H_G11_41' R= -20.45500 -17.02900   3.78400 !END
   !ATOM NAME='H_G12_42' R= -19.64000 -15.46700   4.13200 !END
   !ATOM NAME='H_G13_43' R= -20.19500 -15.83900   2.46500 !END
   !ATOM NAME='H_G21_44' R= -19.27700 -18.79500   2.49700 !END
   !ATOM NAME='H_G22_45' R= -19.04800 -17.63300   1.14700 !END
   !ATOM NAME='H_G23_46' R= -17.63300 -18.49300   1.84000 !END

   !ATOM NAME='H_d47'    R= -15.529   -16.312     1.150   !END

!QM-MM
   !FORCEFIELD
      FF='AMBER'
      PARMFILE='/home/shemmen/PAW/forcefields/amber/cornell_all.prm'
      TOPFILE='/home/shemmen/PAW/forcefields/amber/cornell_all.rtf'
   !END
!END


   !ISOLATE NF=3 RC=0.5 RCFAC=1.5 GMAX2=3.0 DECOUPLE=T !END

!END
!EOB
```

Using 'reduce' to add the hydrogen atoms to the PDB file yields atom labels which do not match the labels in the topology file. Therefore, the user has to correct these labels by hand. Known problems are:

- Hydrogen atoms bonded to the nitrogen atom of the amino acid are labeled as 'H' instead of 'HN'.
- The hydrogen atom of the OH-group in serine has the wrong label.
- Histidine has three different topologies depending on its charge. Possible names are HIP, HIE and HID. The user has to check which residue is the right one and change its name.
- ARG and HID use force field types which are not defined in the original parameter file. They have been added in the current version of the file.

5.4. Visualizing QM/MM simulations

CP-PAW produces a trajectory file in which the position of each atom in each time step is stored. Using the trajectory tool paw_tra a movie can be created to visualize the simulation[13]. However, the trajectory file contains only the QM part of the calculation. In order to see the changes of the environment I let

[13] The trajectory tool can be used for many other aspects. Creating movies is just one of them.

CP-PAW write an additional trajectory file which contains the positions of the atoms of the MM subsystem. The following variables have to be set in the input files:

- Add !CONTROL!ANALYZE!TRA:QMMM=T to your control file. This is set to false by default. If the option is set to true a file called `<rootname>_r.qmmmtra` is generated which contains the coordinate for each MM atom in each time step.
- To create a movie with `paw_tra.x` the user has to make changes in the TCNTL file. At the moment two variable have to be set:
  - !TCNTL:QMMM=T This variable is false by default. If you set it true than the program reads the `_r.qmmmtra`-file instead of the `_r.tra` file. If the file does not exist an error message will be given and the program stops.
  - !TCNTL:FORMAT='PDB' This variable can be switched between 'PDB' and 'STRC'. This determines which input file will be used. QM/MM calculations with the AMBER force field and PDB files as input files require the 'PDB' option.

At the end of the simulation a 'pdb_out' file is also written, which contains the final structure of the MM system.

5.5. Periodic environment

Later in Section 5.6.2 I will show that the simulation of proteins or small peptide chains makes it necessary that the system is solvated in water. Water molecules accept and donate hydrogen bonds to the protein which stabilize its structure. The TIP3P water model is included in the AMBER force field. What is missing is the way the solvent is included into the simulation.

Federmann reports in his diploma thesis the use of water boxes [75]. He provides a generator for water boxes with user defined edge lengths. If a water molecule is going to leave the box during the simulation it bounces off a hard wall. Federmann also implemented a feature to define an ice layer around the water box with an arbitrary thickness. The ice layer is implemented to avoid surface effects. Water molecules, which come close to the edge of the box, still experience the same environment as in the bulk. Seen against this background the radial distribution functions of water have been discussed by Federmann in detail. When simulating systems at finite temperature the water boxes need to be quite large. The simulated molecule of interest must have enough space. Together with the ice layer around the box, the MM systems easily increase to some ten thousand atoms even though the molecule of interest consists of only a few hundred atoms.

In the next step we have implemented periodic MM systems. For we avoid surface effects with periodic systems, the system size can be reduced massively. For example: Consider a water box with an edge length of 25 Å. The ice layer should be at least 5 Å thick so the whole box has an edge length of 35 Å. The

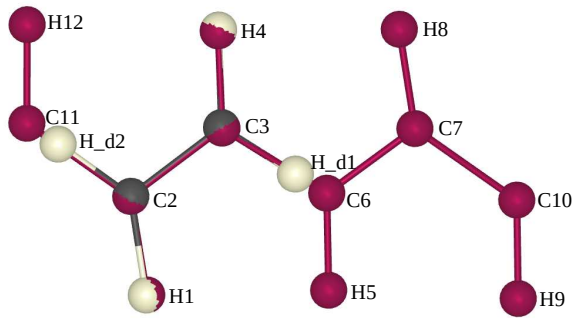


Figure 5.8: The QM and the MM structure of the acetylene unit cell. The purple atoms are atoms of the MM subsystem.

inner part of the box has only 36% of the volume of the system. For a larger box with 100 Å edge length, the inner part is still 75% of the total volume. Using periodic MM systems surface effects can be neglected and a lot of computing time can be saved.

Another advantage of periodic MM systems is that we are able to simulate periodic structures. Our first test system has been an acetylene chain. The challenge in periodic systems is to describe bonds which cross the cell boundary. Figure 5.8 shows the MM structure of the unit cell and the QM system. The acetylene chain has been simulated using UFF, for acetylene is not parameterized in the AMBER topology file. The MM unit cell is given in the structure file by:

```
1  !STRUCTURE
2  ....
3   !QM-MM
4     !ATOM NAME='H_1'   FFTYPE='H_'  QMATOM='H_1'  !END
5     !ATOM NAME='C_2'   FFTYPE='C_2' QMATOM='C_2'  !END
6     !ATOM NAME='C_3'   FFTYPE='C_2' QMATOM='C_3'  !END
7     !ATOM NAME='H_4'   FFTYPE='H_'  QMATOM='H_4'  !END
8     !ATOM NAME='H_5'   R= 2.5   0.5   0.0  FFTYPE='H_'   !END
9     !ATOM NAME='C_6'   R= 2.5   1.5   0.0  FFTYPE='C_2'  !END
10    !ATOM NAME='C_7'   R= 3.5   2.4   0.0  FFTYPE='C_2'  !END
11    !ATOM NAME='H_8'   R= 3.5   3.4   0.0  FFTYPE='H_'   !END
12    !ATOM NAME='H_9'   R= 4.5   0.5   0.0  FFTYPE='H_'   !END
13    !ATOM NAME='C_10'  R= 4.5   1.5   0.0  FFTYPE='C_2'  !END
14    !ATOM NAME='C_11'  R=-0.5   2.4   0.0  FFTYPE='C_2'  !END
15    !ATOM NAME='H_12'  R=-0.5   3.4   0.0  FFTYPE='H_'   !END
16
17    !LINK MMJOINT='C_2' MMATOM='C_11' QMATOM='H_d1' SHFFTYPE='H_' !
          END
18    !LINK MMJOINT='C_3' MMATOM='C_6'  QMATOM='H_d2' SHFFTYPE='H_' !
          END
19
20    !BOND ATOM1='H_12' ATOM2='C_11' BO=1.0   !END
21    !BOND ATOM1='C_11' ATOM2='C_2'  BO=1.0   !END
22    !BOND ATOM1='H_1'  ATOM2='C_2'  BO=1.0   !END
23    !BOND ATOM1='C_2'  ATOM2='C_3'  BO=2.0   !END
24    !BOND ATOM1='C_3'  ATOM2='H_4'  BO=1.0   !END
25    !BOND ATOM1='C_3'  ATOM2='C_6'  BO=1.0   !END
```

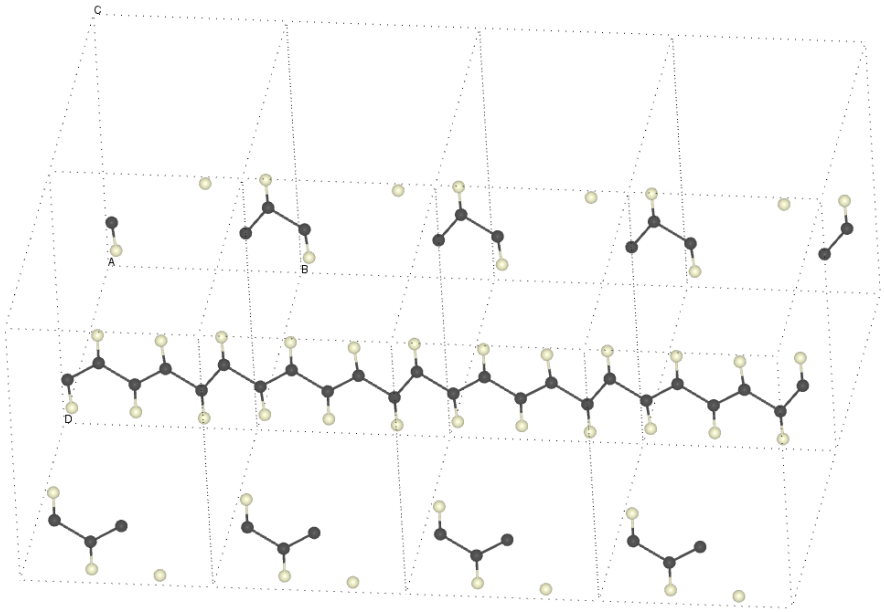


Figure 5.9: Acetylene chain in a 4x1x2 cell.

```
26     !BOND ATOM1='C_6'   ATOM2='H_5'   BO=1.0   !END
27     !BOND ATOM1='C_6'   ATOM2='C_7'   BO=2.0   !END
28     !BOND ATOM1='C_7'   ATOM2='H_8'   BO=1.0   !END
29     !BOND ATOM1='C_7'   ATOM2='C_10'  BO=1.0   !END
30     !BOND ATOM1='H_9'   ATOM2='C_10'  BO=1.0   !END
31     !BOND ATOM1='C_10'  ATOM2='C_11:+1+0+0' BO=2.0  !END
32
33     !LATTICE T[A]= 6.5  0.  0.  0.  1000.  0.  0.  0.  1000.  !END
34   !END
35 !END
36 !EOB
```

In line 33 the unit cell of the MM system is defined. The lattice parameter in x-direction has to be optimized for the system. The y- and z-direction have been chosen large enough so that the neighboring cells do not interact. As it can be seen in Figure 5.8 atom number ten has a bond to atom number eleven in the next unit cell. This is set in line 31. We have introduced an extended notation for bonds crossing the cell boundary. The atom is defined by its name followed by a colon and the translation vector. In this case, the translation vector is (1,0,0). Figure 5.9 shows the resulting structure of the acetylene chain in a 4x1x2 cell[14].

We will now take a closer look at the implementation of periodic MM systems in the CP-PAW program package. The major challenge is to construct an accurate neighbor list. But before the neighbor list is set, the exclusions have to be evaluated. Non-bond interactions such as van-der-Waals and Coulomb interac-

[14] For a better illustration the lattice vectors have been set to (6.5, 0, 0), (0, 10.0 , 0) and (0, 0, 10.0) in CryMolCAD.

tions must not be calculated for pairs of atoms that are connected via a bond or by a common atom. In the case of AMBER calculations the 1-4 bonded atoms are added to the exclusions, too, because the van-der-Waals and the Coulomb interaction have to be scaled.

The exclusions are a list of those neighbors. One exclusion contains the information of the two atoms in the pair and the integer translations of the second atom. In order to reduce memory, all five numbers are mapped onto a single integer number. The non-bond exclusions are stored as a one-dimensional array:

$$1+(IT1+1)+3*((IT2+1)+3*(IT3+1)+3*((IAT1-1)+NAT*(IAT2-1))) \tag{5.5.1}$$

IAT1 and IAT2 are the index numbers of the two atoms. NAT is the number of all atoms. IT1, IT2 and IT3 are the integer translations in each direction. The program uses the convention that IAT1 is always greater than IAT2. At the end the exclusion array is sorted by increasing magnitude. An additional array I2FIRST is created. This array assigns each atom the first position where it occurs in the exclusion list. When the program looks for a special exclusion it does not have to search the whole exclusion list. It only has to search the I2FIRST array which gives the starting point in the exclusion list.

The subroutine CLASSICAL_NEIGHBORS sets up the neighbor list. The neighbor list is an array (NBLIST) which is of type NONBOND_TYPE:

```
1  TYPE NONBOND_TYPE                    ! NONBOND NEIGBORLIST ITEM
2    INTEGER(4)            :: IAT1      ! INDEX OF FIRST ATOM
3    INTEGER(4)            :: IAT2      ! INDEX OF SECOND ATOM
4    INTEGER(4),POINTER    :: IT(:)     ! INTEGER LATTICE TRANSLATION OF
         SECOND ATOM
5    LOGICAL(4)            :: EXCLUDE   ! BOND EXCLUSION
6    LOGICAL(4)            :: TONEFOUR
7  END TYPE NONBOND_TYPE
```

The subroutine divides the unit cell into smaller boxes[15] and attributes each atom to a sub-box. The variable NATINBOX holds the information how many atoms are part of the box. Other variables needed to be set are FIRST which points to the first atom in a given box and IATPNT which points from the atoms in a given box to the bare atom index.

To build the neighbor list a loop runs over all atoms. For each atom the program looks for the atoms in the neighboring boxes. If the distance between two atoms is less than RCLONGRANGE[16] then the atom pair is defined to be neighbors. The program creates an entry in the neighbor list which contains the atom pair and the integer translation vector of the second atom.

The final variable which has to be set is NBLIST%EXCLUDE. Therefore, the atom pair is translated into an integer number as shown in Equation (5.5.1). This number has to be compared with the exclusion list. At this point the program can benefit from the I2FIRST array. If an exclusion exists which is equal to the atom pair in the neighbor list, the variable EXCLUDE is set to false. The variable TONEFOUR is taken from the exclusion list and determines

[15] In the current implementation it divides the unit cell into 3x3x3=27 sub-boxes.
[16] RCLONGRANGE is actually set to 100 a_0.

whether van-der-Waals and Coulomb interactions have to be scaled.

5.6. Applications

5.6.1. The water dimer

The first system that has been simulated with the AMBER force field is the water dimer. The water dimer has been studied experimentally and theoretically [90, 91, 92, 72, 93] and its structure is well known. Therefore, it is a good test for the water model in AMBER.

Two water molecules placed arbitrarily define the start structure. The AMBER_MD program has been used for the calculation. I have simulated 5000 time steps with 0.25 fs each time step. A constant friction of 0.001 has been applied. The system has been placed in a cubic cell with 800Å side length. Table 5.1 shows the coordinates of the atoms of the water dimer before and after the simulation. The same system has been calculated as a pure QM sim-

		start structure			final structure		
	Atom	x	y	z	x	y	z
Molecule 1	O	0.000	0.000	-2.000	-0.049	0.075	-1.320
	H	0.000	1.000	-2.000	0.492	-0.445	-1.913
	H	-1.000	0.000	-2.000	-0.831	0.285	-1.831
Molecule 2	O	0.000	0.000	2.000	-0.013	-0.034	1.424
	H	-0.300	0.700	2.000	0.495	0.752	1.625
	H	0.400	-0.500	2.000	-0.061	-0.051	0.468

Table 5.1: Start and final structure of the water dimer calculated with the AMBER_MD program.

ulation with the CP-PAW program. The lattice vectors of the unit cell are $T_1 = (0.0, 7.0, 7.0)$, $T_2 = (7.0, 0.0, 7.0)$ and $T_3 = (7.0, 7.0, 0.0)$. The structural properties of the two simulations compared with experimental data is given in Table 5.2. α is the so-called donor angle. It is the angle between the bond building hydrogen atom, the oxygen atom from the same molecule and the oxygen atom from the other molecule. β is the acceptor angle. Figure 5.10 shows the dimer and the labels.

The results of both calculations are satisfactory. The results from the PAW calculation agree with the experimental data in all cases. The results from the

	AMBER	CP-PAW	Experiment
R_{O-O} in Å	2.7464 (- 7%)	2.8196 (-5%)	2.9760
α in degree	3.92	6.53	6 ± 20
β in degree	17.35	51.70	57 ± 10

Table 5.2: Structural data of the water dimer. The experimental data is taken from [92].

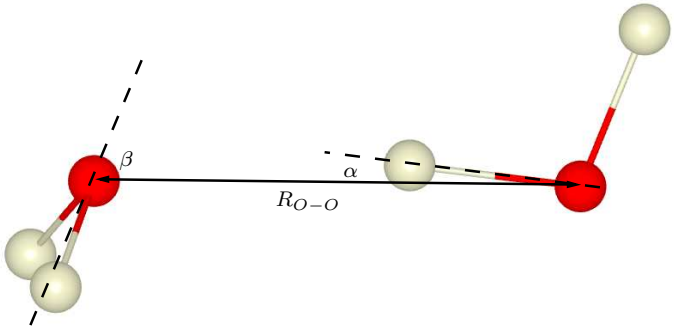


Figure 5.10: The water dimer.

AMBER calculation agree for the O-O distance and for the donor angle. The acceptor angle is too small. We can accept this result at the moment because the main goal is that the first solvation shell around a protein is described correctly. While the simulation shows a correct hydrogen bond between the water dimer, the water molecules in simulations with proteins are able to stabilize the protein structure. This will be discussed in the next section. Looking at the radial (RDF) distribution function of TIP3P water simulations one can see that the second peak in the RDF for the O-O distance is missing [62]. To get a better solvation model the implementation of the TIP4P is planned.

5.6.2. Small peptide chains

The next test systems have been the tri-peptide introduced in Section 5.3 and the 1AMB structure. The need of a surrounding water environment can be demonstrated on 1AMB and on the tri-peptide system. Without the aqueous environment the protein's secondary structure is lost. Figure 5.11 shows the peptide in different surroundings. In vacuum the peptide looses its initial structure. Intramolecular hydrogen bonds bring both ends of the molecule closer together. Adding water as solvent it can be seen that water molecules fill the space between both ends of the peptide. The water molecules donate or accept hydrogen bonds which stabilizes the structure of the molecule. This shows the importance of a water model in the force field and QM/MM simulations.

The 1AMB structure contains 28 residues of the amyloid β-peptide. The β-peptide is the major proteinaceous component of amyloid deposits in Alzheimer's disease. It has been reported by Talafous et. al. [94]. In membranlike media, the peptide folds to form a predominately α-helical structure with a bend centered at residue 12. This α-helical structure has been the main reason to choose this system for testing the AMBER implementation. On the one hand I could test how the force field works with a larger structure[17]. On the other hand it has been an interesting simulation which has shown the need and the effect of an aqueous environment to conserve the α-helical structure.

[17] Note that 1AMB is still a really small structure compared to most other proteins.

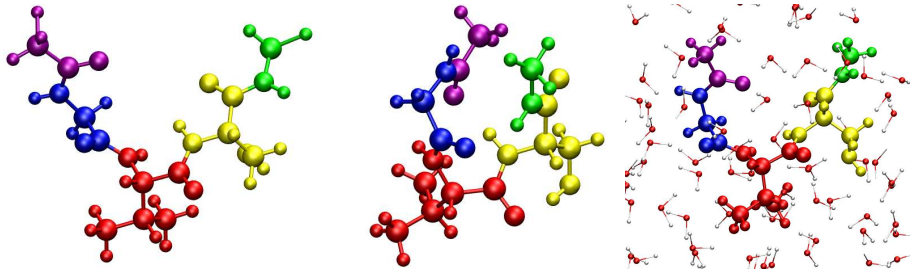


Figure 5.11: The tri-peptide structure. Different colors label different amino acids. Left: the tri-peptide in its initial form from the PDB file. Middle: the tri-peptide after geometry optimization in vacuum. The peptide is not stabilized by a surrounding solvent. Thus, it looses its initial form. Both ends come closer, intramolecular hydrogen bonds are formed. Right: the tri-peptide in aqueous solution. Water molecules stabilize the structure by offering hydrogen bonds between both ends of the molecule.

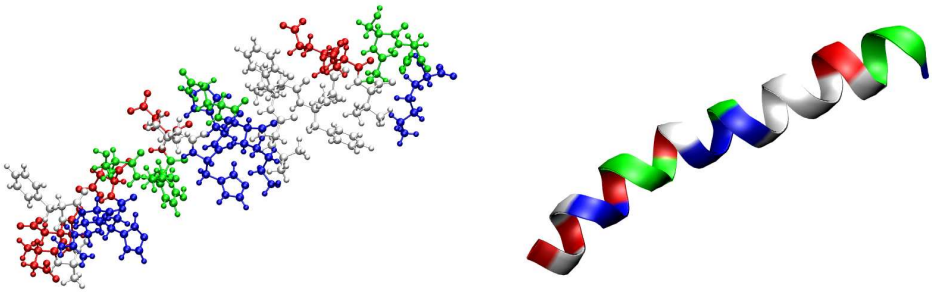


Figure 5.12: The 1AMB structure from the protein data bank. In the left part the atoms are drawn. Different colors label different amino acids. In the right part the α-helix is shown.

Figure 5.12 shows the 1AMB sequence from the PDB file. The same behavior can be found at the 1AMB structure. A simulation without solvent lets the molecule loose its α-helical structure. This can be seen in Figure 5.13. Adding water into the simulation cell the α-helical structure is conserved during the simulation. This can be seen in Figure 5.14.

The final structure which has been tested is the Fe_4S_4 cluster of the Fe-protein in Nitrogenase[18]. The cluster has four links to the protein environment. Each link is followed by six amino acids in the test system. The Fe_4S_4 cluster is known in three different charge states. All three states have been parameterized in the topology file. Therefore, CP-PAW calculations have been done for each state. The bond lengths and charges have been transferred to the topology file. Table 5.3 shows the topologies.

The cluster has been simulated with AMBER and in a QM/MM simulation. The system has been surrounded by water molecules to ensure stabilization of the protein structure.Figure 5.15 shows the system after calculated in a QM/MM

[18] The Fe-protein is introduced in Section 6.1.

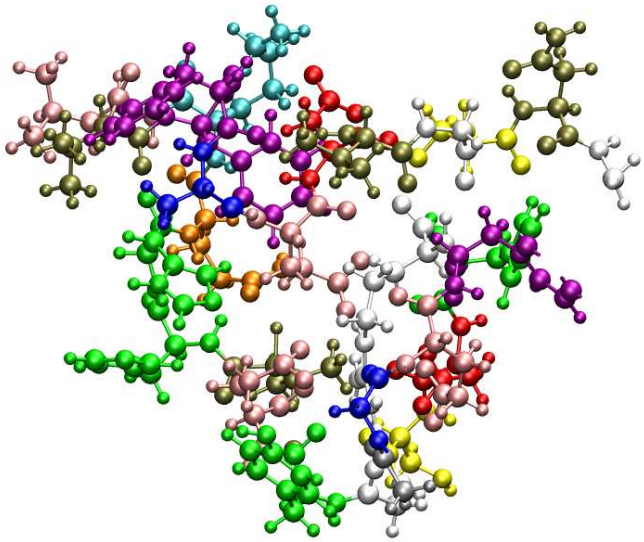


Figure 5.13: The 1AMB structure optimized in vacuum. The α-helical structure is completely lost.

simulation. The water molecules are not shown in the figure. The green residues are the cysteines which are linked to the cluster itself.

5.7. Conclusions and Outlook

In this Chapter the implementation of the AMBER force field in the CP-PAW program package has been presented. Some applications of the AMBER force field have been shown, too.

The last application has been the Fe_4S_4 cluster of the Fe-protein of Nitrogenase. The cluster itself is not part of the original AMBER topology file. I have reported a simple way to parameterize such a structure. The parameterization does not have to be very accurate, for the error introduced by a bad parameterization is subtracted out in the QM/MM framework.

The next step should be to parameterize the other iron-sulfur clusters of Nitrogenase. This will allow simulations of the protein which can lead to a deeper understanding of the proton channels to the FeMo-cofactor.

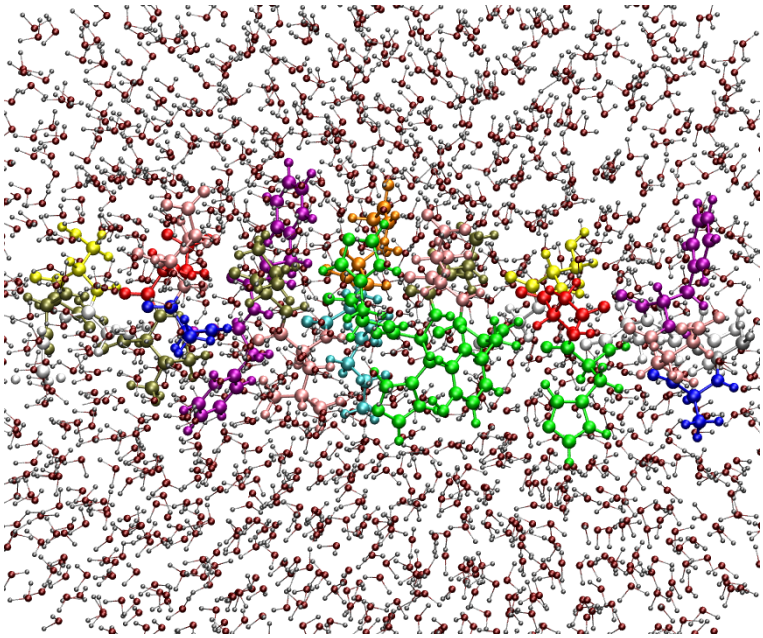


Figure 5.14: The 1AMB structure optimized in water. The α-helical structure is conserved.

		Charge		
atom	force field type	$[Fe_4S_4]$	$[Fe_4S_4]^{+}$	$[Fe_4S_4]^{2+}$
FE1	F1	0.58962	0.55544	0.50338
FE2	F1	0.50238	0.50385	0.47933
FE3	F1	0.67589	0.60374	0.52426
FE4	F1	0.62129	0.53793	0.47595
S1	SF	-0.49360	-0.47193	-0.43098
S2	SF	-0.53668	-0.53362	-0.48505
S3	SF	-0.50046	-0.46561	-0.40951
S4	SF	-0.47483	-0.44395	-0.39147
residue name		F40	F41	F42

bond pairs					
FE1 - S1	FE1 - S2	FE1 - S3	FE2 - S1	FE2 - S2	FE2 - S4
FE4 - S2	FE3 - S3	FE3 - S4	SE4 - S1	FE4 - S2	FE4 - S4

Table 5.3: Atom names, force field types and charges of the Fe_4S_4 cluster in the topology file. The second table shows the bond pairs in the cluster.

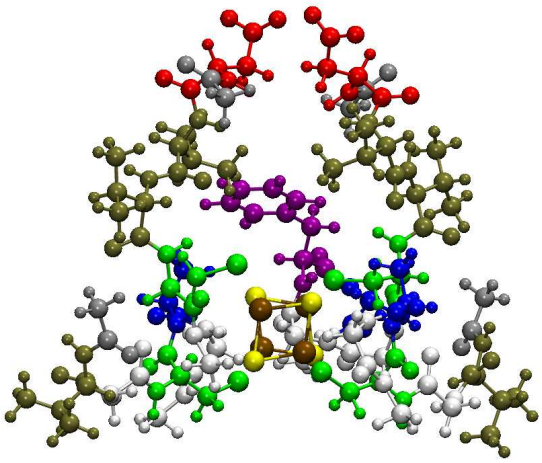


Figure 5.15: The Fe_4S_4 cluster of the Fe-protein of Nitrogenase. The cluster itself is the QM-part of a QM/MM simulation using the AMBER force field. The green amino acids are the cysteine residues which embed the cluster into the protein.

6. Structural changes of the P-cluster

6.1. Biological Nitrogen Fixation - a Review

Examining biological systems on the atomic scale reveals only five major elements. These elements are hydrogen, oxygen, phosphorus, carbon and nitrogen. However, in nature only hydrogen, oxygen and phosphorus exist in a sufficient amount in a metabolic form. The primary source of metabolic usable carbon in nature is carbon dioxide (CO_2),the source of nitrogen is the atmospheric dinitrogen (N_2). CO_2 is processed by plants in a photosynthetic reaction to a metabolic form. Dinitrogen can only be transformed by some special strains of bacteria into a metabolic form. Thus, nature suffers from a lack of nitrogen, even though nearly eighty per cent of the atmosphere consists of nitrogen. The reason why the atmospheric nitrogen is inaccessible is the inert form of the dinitrogen. Figure 6.1 shows the molecular energy levels of a dinitrogen molecule. All bond

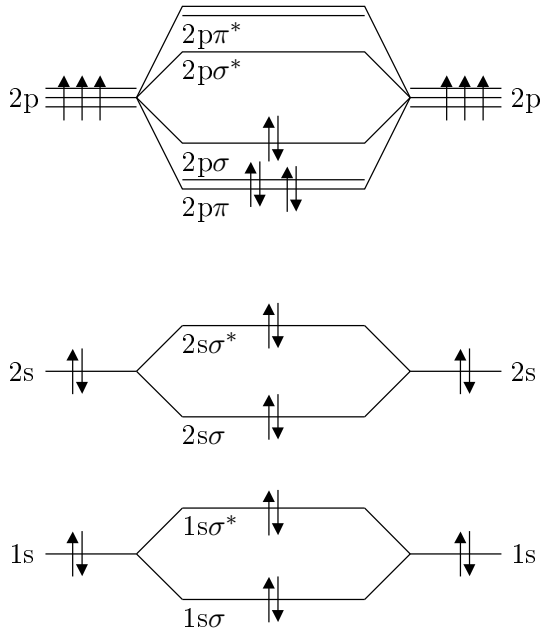


Figure 6.1: The schematic molecular orbitals of dinitrogen

building orbitals (2pσ and 2pπ) are completely filled with electrons. This explains the triple bond of dinitrogen and its inert behavior. Adding two more electrons to the system (O_2 molecule) destabilizes the chemical bond because the anti-bonding 2pσ* molecular orbital will be filled.

There are only two ways to bring the atmospheric nitrogen into the biological cycle: The energy of lightning is high enough to break the dinitrogen bond. This results in special nitrogen salts in the soil. But this nitrogen source is not able to provide the amount of nitrogen which is needed in nature. The second possibility is the ability of special bacteria to break the dinitrogen bond. They use a special enzyme called Nitrogenase to transform dinitrogen into ammonia.

The catalytic reaction was first described in the Thorneley-Lowe scheme [95, 96, 97, 98]. The netto reaction is

$$N_2 + 8H^+ + 8e^- + 16\text{MgATP} \longrightarrow 2NH_3 + H_2 + 16\text{MgADP} + 16P_i$$

which is exergonic under room temperature with $\Delta G = -15.2\text{kcal mol}^{-1}$ [99]. Furthermore, the hydrolysis of MgATP provides additional 112kcal mol^{-1} [100, 101]. It is possible that one catalytic cycle produces only H_2 instead of N_2. The mechanism of the hydrogen production and the control of it is still discussed in literature. Kästner proposed that molecular hydrogen is produced if one of the iron atoms of the FeMo-cofactor is protonated. The hydrogen atom becomes a hydride (H^-) which forms a H_2 molecule together with an additional proton [102]. Even under high partial pressure of dinitrogen the hydrogen production occurs [103, 104].

The enzyme Nitrogenase consists of two proteins [105, 106]. One is the Fe-protein. Its role in the catalytic process is to provide electrons for the reaction. It hydrolyzes MgATP, thus delivering two electrons in each step [107]. The electrons are transported via an iron-sulfur cluster from the Fe-protein to the other protein when the Fe-protein docks at the latter, the so-called MoFe-protein. The MoFe-protein itself contains two iron-sulfur clusters. The P-cluster takes the electrons from the Fe-protein and delivers them to the FeMo-cofactor, the second iron-sulfur cluster, where the nitrogen fixation occurs. Figure 6.2 shows the enzyme[1]. It is based on the data from the 1M1Y structure [108].

A more detailed view of the MoFe-protein is given in Figure 6.3. The figure has been created from the data of the 1M1N protein structure. The 1M1N structure has a resolution of 1.16Å. It was published by Einsle in 2002 [109]. Einsle was the first who was able to prove the existence of the central ligand in the FeMo-cofactor. Figure 6.4 shows the FeMo-cofactor and the P-cluster. The P-cluster is depicted in two oxidation states. This will be discussed later.

The main interest of recent research has been the FeMo-cofactor and the mechanism of the catalytic cycle of the nitrogen reduction. Crystallographic experiments [110, 109] have shown that the FeMo-cofactor is a $MoFe_7S_9N$-cluster. The iron atoms of the cluster form two tetrahedrals which are bridged over a surface side. The discovery of the central ligand in 2002 changed the basis of all former theoretical calculations [111, 112]. While the existence of the central ligand has been shown by Einsle et. al. [109] it is still discussed whether it is a nitrogen atom or not [113, 114].

Three different models of the catalytic mechanism have been suggested which are based on DFT calculations and include the central ligand. There are two

[1] It is only one half of the enzyme. The whole Nitrogenase is a complex of two Fe-proteins and two MoFe-proteins.

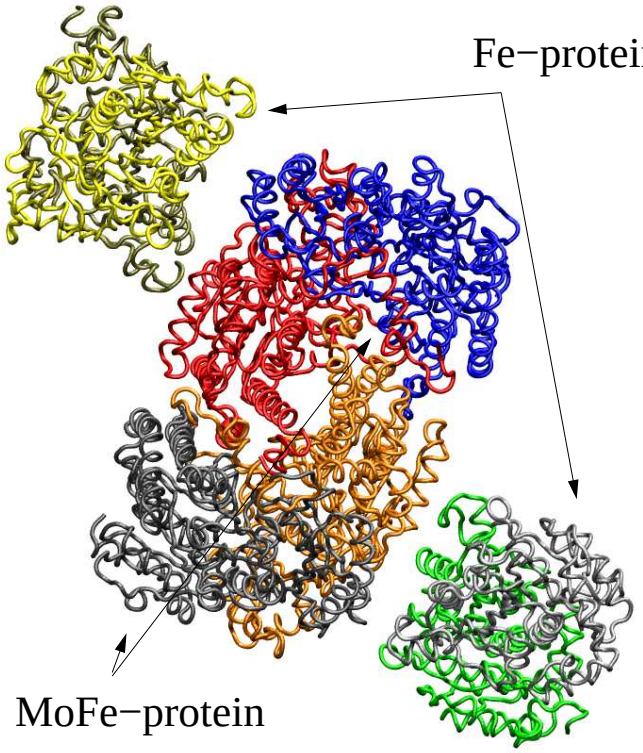


Figure 6.2: The enzyme Nitrogenase with its two subunits, the Fe-protein and the MoFe-protein. Different colors label different protein chains.

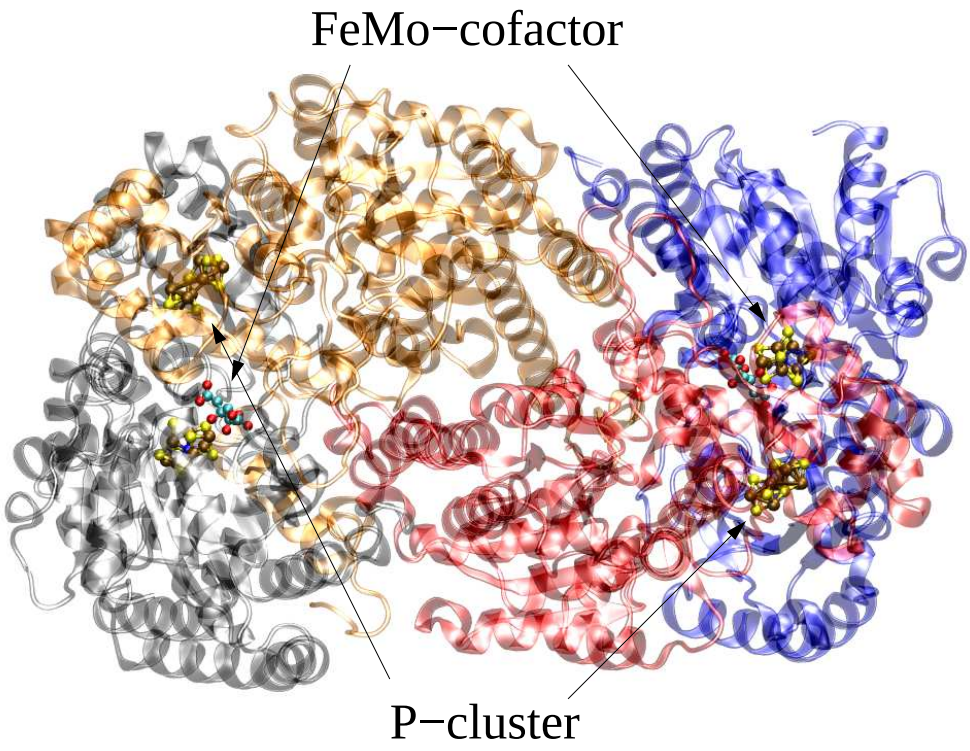


Figure 6.3: The MoFe-protein from the 1M1N structure [109]. The P-cluster and the FeMo-cofactor are pointed out.

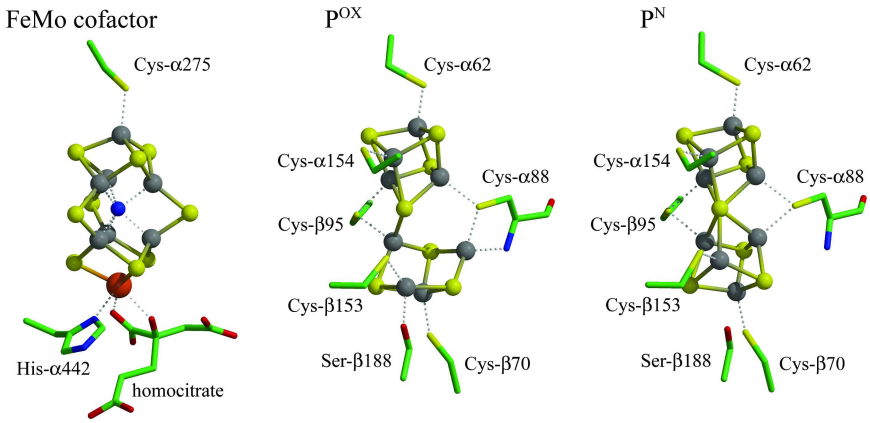


Figure 6.4: The iron-sulfur clusters of the MoFe-protein of Nitrogenase. On the left side, the FeMo-cofactor is shown On the right side the P-cluster is depicted in its resting state. In the middle is the P-cluster in the oxidation state. The pictures are printed with kind permission from Prof. Einsle

additional models which shall be reported here for completeness, which are more important in the view of model complexes than in their contribution to the explanation of the biological nitrogen fixation.

Hinnemann et. al. [115] expanded their earlier results [116, 117] to include the central ligand of the FeMo-cofactor. In their model the dinitrogen binds to one of the iron atoms of the central prismatic structure of the cofactor. The dinitrogen molecule is then reduced and protonated gradually. In their work all intermediates have been investigated. The binding energy of the dinitrogen to the cluster is reported to be -60kJmol^{-1} which is not favorable. Calculations done in our group predict a stable bond between dinitrogen and the cofactor only after breaking a sulfur bridge of the cofactor.

Another model has been proposed by Huniar et. al. [118]. The first step in their model is the reduction of the central ligand to ammonia. In the following step dinitrogen binds to a iron atom, which results in the breaking of a protonated sulfur bridge. This opening of the cage after protonation of a sulfur atom corresponds to the model developed by Kästner [102]. After adding three protons and electrons to the dinitrogen molecule, ammonia separates from the molecule. The remaining nitrogen atom replaces the central ligand and the cycle can start again. But this model stands in contrast to experiments which prove that the central ligand is not replaced during the reaction cycle [113]. It has to be mentioned at this point that the discussion about the interpretation of these experiments is not finished.

The third model has been developed in our group by Kästner et. al.[102, 119]. It proposes that one SH-bridge opens when the dinitrogen molecule binds to the cluster. This allows the dinitrogen to bind as a bridge between two iron atoms. These two weak bonds of the dinitrogen activate the molecule to make it accessible for the first protonation. After repeated reduction and protonation steps the two iron atoms are only bridged by one nitrogen atom. The first ammonia molecule leaves the cluster after adding another proton. In the next step the hydrogen atom from the sulfur bridge protonates the nitrogen atom. Ammonia leaves the cluster and the sulfur bridge is closed again. The cluster has returned to its original conformation and the catalytic cycle is completed. Figure 6.5 shows the complete cycle. In the proposed mechanism the central ligand of the FeMo-cofactor plays an important role. The central atom is able to offer bonds to the iron atoms. In all association and dissociation steps during the catalytic cycle it ensures that the iron atoms of the cluster remain in a tetrahedral configuration. The tetrahedral configuration is the preferred configuration of the iron atoms. Another important factor is the opening of the sulfur bridge after protonation. This allows the dinitrogen molecule to bind in a bridging mode between two iron atoms of the cluster.

The first protonation of dinitrogen is the most difficult step in the catalytic cycle. In the gas phase the reaction $N_2 + H^+ + e^- \longrightarrow N_2H$ is endothermic and requires the energy of 164kJmol^{-1}. The main goal of Nitrogenase is to lower this barrier. In the bridged mode the N-N distance of dinitrogen is elongated compared to the distance in the gas phase as well as the vibration frequency is reduced. This indicates the activation of the triple bond for the following protonation. The activation and protonation of dinitrogen has been studied

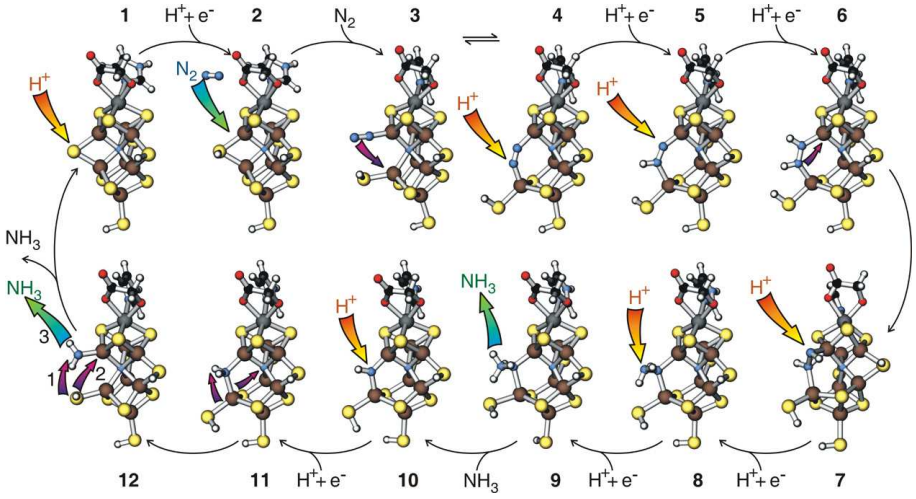


Figure 6.5: The complete reaction mechanism of the nitrogen fixation in Nitrogenase. Printed with kind permission from Dr. Kästner.

in detail in our group. The results have been reported in [120]. As shown in [102], dinitrogen can dock at the FeMo-cofactor in two configurations with similar energies: the axial binding mode, denoted by $A0$, is slightly more stable than the bridged one, denoted by $A1$. The activation of the triple bond can be traced back to occupied N_2-π^* orbitals as shown in Figure 6.6. These orbitals can be seen as an antisymmetric combination of the Fe-N bonds. Interestingly they are only dominant in the minority spin direction of the two neighboring Fe sites. The origin is that the interaction with the unoccupied d states split the π^* orbital in a bonding and an anti bonding orbital. The bonding orbital, still having π^* character but containing the Fe-N bonding contribution, becomes occupied and is located about 1 eV below the HOMO, while the anti bonding orbital, having the Fe-N anti bonding contribution, remains unoccupied. The former, bonding orbital is the relevant frontier orbital for the protonation.

The opening of the cage together with a bridging dinitrogen molecule between iron atoms has also been reported by Sellmann et. al. [121, 122, 123, 124]. They showed the analogy of this mechanism to smaller iron complexes. The difference between their models and the FeMo-cofactor is the coordination number of the iron atoms. In the model reported by Kästner the iron atoms form tetrahedral high-spin complexes while Sellmann's reported models form octahedral low-spin complexes. Taking this idea Reiher et. al. investigated the reaction with organometallic iron- and ruthenium-complexes. They proposed an excitation of the triple bond by photochemical effects in order to activate the bridging dinitrogen molecule.

Not only the iron atoms of the FeMo-cofactor have been investigated and discussed as possible docking stations for the dinitrogen. A number of theoretical works exist which favor the molybdenum atom to form the first bond with the nitrogen [125, 126, 127, 128, 129]. Recently a catalytic reaction cycle has been reported which shows the nitrogen fixation by a molybdenum complex [130]. This seems to strengthen the role of the molybdenum atom in the FeMo-

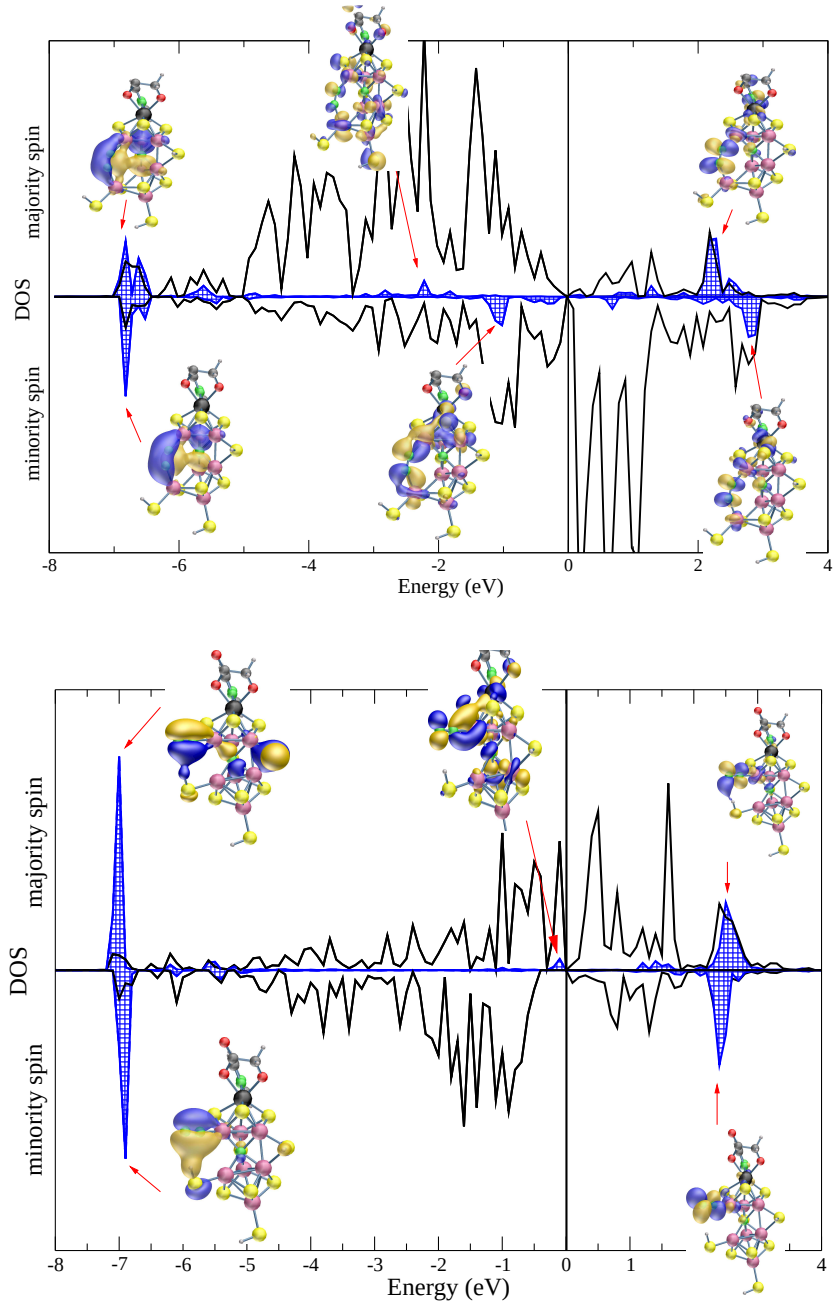


Figure 6.6: Density of states of the bridged ($A1$, top) and the axial ($A0$, bottom) binding modes illustrating N_2 activation. The full line indicates the density of states projected onto the d orbitals of the Fe atoms directly bound to N_2. the shaded curve is the density of states projected onto the π and π^* orbitals of dinitrogen. The insets show the relevant wave functions. Note that both states have occupied π^* orbitals, which is more dominant and lower in energy in the bridged binding mode $A1$.

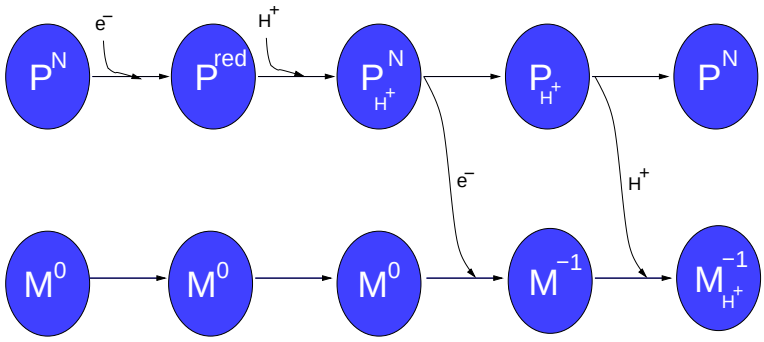


Figure 6.7: Proposed proton-electron mechanism between P-cluster (P) and FeMo-cofactor (M).

cofactor. However, the calculations done by Kästner yield a total energy which is 50kJmol^{-1} higher in the case of docking at the molybdenum atom than docking to a iron atom. Thus, mechanisms where nitrogen attacks the molybdenum atom are not compatible with the mechanism developed in our group.

6.2. The role of the P-cluster

6.2.1. An Overview

The transfer of electrons from the Fe-protein to the MoFe-protein has been studied by Kurnikov et. al. [100]. The Fe_4S_4 cluster lying at the surface of the Fe-protein experiences a change of the environment while the Fe-protein docks at the MoFe-protein. Originally lying in an aqueous environment the cluster comes into a protein environment after docking. The Fe_4S_4 cluster is no longer shielded by the dielectric solvent. The electron becomes destabilized and is transported to the P-cluster inside the MoFe-protein. Hydrolization of ATP initializes the separation of the Fe-protein from the MoFe-protein.

The P-cluster transfers the electrons needed by the nitrogen fixation to the FeMo-cofactor. The P-cluster is a Fe_8S_7 iron-sulfur cluster. Its structure has finally been revealed in 1997 by comparing the crystal structure of the oxidized and the unoxidized state [131, 132]. The structure of the resting state, called P^N, is shown in the right part in Figure 6.4. The cluster is built by two Fe_4S_4 cubanes which are connected by a common sulfur atom. All iron atoms are surrounded by four sulfur atoms (some of them are part of the protein environment), thus forming tetrahedral complexes.

Taking a closer look at the oxidized state of the P-cluster, P^{ox}, shows that the cluster builds two additional bonds to the protein in this state (central part of Figure 6.4). These conformational changes are the basis of many speculations of the role of the P-cluster. It has been discussed which influences the conformational changes on the protein have. One possibility is that the proton and electron transfer through the protein can be more controlled in this way. This would lead to a possible control of the H_2 production. Another possibility would be the control of the association and dissociation of the Fe-protein to

the MoFe-protein [132]. It has to be remarked that the reduced state of the P-cluster is involved in nitrogen fixation and not the oxidized state. Figure 6.7 shows my proposed mechanism of the electron and proton transfer to the FeMo-cofactor. The reduced state of the P-cluster is involved in the electron transfer to the FeMo-cofactor. Unfortunately, the reduced state has not been identified in experiment yet. If the P-cluster gets an electron from the Fe-protein, the question arises, where the electron can be placed at the P-cluster. In the resting state, all iron atoms are in a Fe^{2+} state. Adding one electron a Fe^{1+} would be created, which would be an unphysical state. The sulfur atoms are also a bad choice for placing the electron, for they are already in a S^{2-} state.

The oxidized state is of interest because experimentalists have access to it and detailed experimental data of its structure and its spin signals have been reported [133, 110, 134, 135, 131, 136].

The resting state of the P-cluster produces no EPR signal, i.e. $S = 0$. The twofold oxidized state P^{ox} produces $S = 3$ or $S = 4$ [133]. The EPR signals observed for the one-fold oxidized state $P^{semi-ox}$ show an overlay of $S = \frac{1}{2}$ and $S = \frac{5}{2}$ [137, 138, 107]. This implies two alternative interactions with the protein environment. The P-cluster has been studied in 1994 with theoretical methods [135]. That work focused on the spin states and the hyperfine parameters of the P-cluster. However, the model of the P-cluster used in this work has been a Fe_8S_8 cluster, which is incorrect.

6.2.2. The input structure

The basis of the structure for all reported calculations in this thesis is the PDB file **3MIN** [131] from the protein data bank [139]. 3MIN contains the structure of the MoFe-protein from Azotobacter vinelandii which has been refined to 2.0Å resolution. The best structure of the MoFe-protein available is the 1M1N structure [109]. It contains the Fe-protein and the MoFe-protein refined to 1.16Å. For the structure of the P-cluster is identical in both files, the 3MIN structure has been used because of its smaller size. The oxidized form of the P-cluster can be found in the 2MIN structure [131].

A model of the P-cluster together with some atoms of the environment is shown in Figure 6.8. This model is used as input structure for the calculations. The input structure contains the Fe_8S_7 iron-sulfur cluster. Assuming that each iron atom is in a ferrous Fe^{2+} state [133] and each sulfur atom is S^{2-}, the total charge of the bare cluster is 2+. The P^N structure has six links to the protein. Each link is built by a cysteine residue. The amino acid cysteine is characterized by a CH_3SH group. Without the hydrogen atom the cysteine builds a bond to a nearby iron atom of the P-cluster. Each sulfur atom provides a negative charge which sums up to a total charge of minus four for the whole system. Most cysteine residues have been cut after the first atom which is not directly connected to the cluster itself. The cut atoms are saturated with hydrogen atoms. It is known that the nitrogen atom of the residue Cys-β153 builds a bond to the P-cluster in its oxidized state. Therefore the residue is not cut after the first atom. For symmetry reasons the backbone of residue Cys-α88 is also included into the simulation cell. The part of the serine residue is also part of

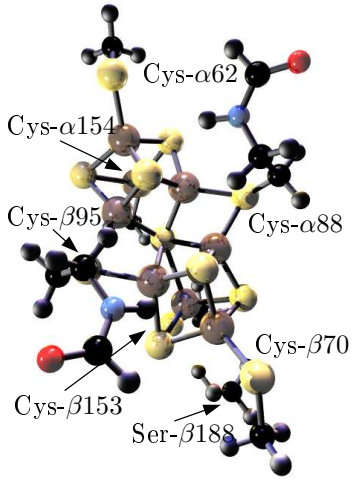


Figure 6.8: P-cluster P^N from the 3MIN PDB file with some additional atoms from the protein environment. This structure is used as basic input structure in the calculations reported in this thesis.

the cell because it will build a bond to P^{ox}. More details of the calculation can be found in Appendix A. The structure file for the P^N structure is printed in Appendix B.

6.2.3. Proton pathways

In this section I will present a possible pathway for protons from the P-cluster to the FeMo-cofactor.

Comparing the 3MIN and 2MIN structure one can see that there are two hydrogen atoms less in the oxidized system. The oxygen atom from serine discards the proton and forms a bond to the nearby iron atom. Also the nitrogen atom from cysteine loses its hydrogen atom and builds a bond to the P-cluster. A question of interest is where these hydrogen atoms will be found and where they come from when the P-cluster gets back into the resting state.

Figure 6.9 shows the P-cluster and the serine-188 residue. The 1M1N structure has been used for the following considerations, for it provides the most accurate X-ray structure including water molecules. Serine-188 lies near the protein surface and is surrounded by water molecules. The water molecules are able to transport protons from the aqueous environment of the protein to the P-cluster.

I found a possible pathway for the proton from the cysteine residue. Following that path the proton can reach the FeMo-cofactor. To build the proton path I searched for water molecules or parts of amino acids which can form hydrogen bonds. Table 6.1 lists the possible residues which can be part of a proton path. The path from the P-cluster to the FeMo-cofactor is drafted in Table 6.2. Figure 6.10 shows the the path in the protein.

The proton path is important for the interaction between P-cluster and FeMo-cofactor. The FeMo-cofactor needs electrons and protons during the nitrogen fixation. While it is already known that the electrons are provided by the P-

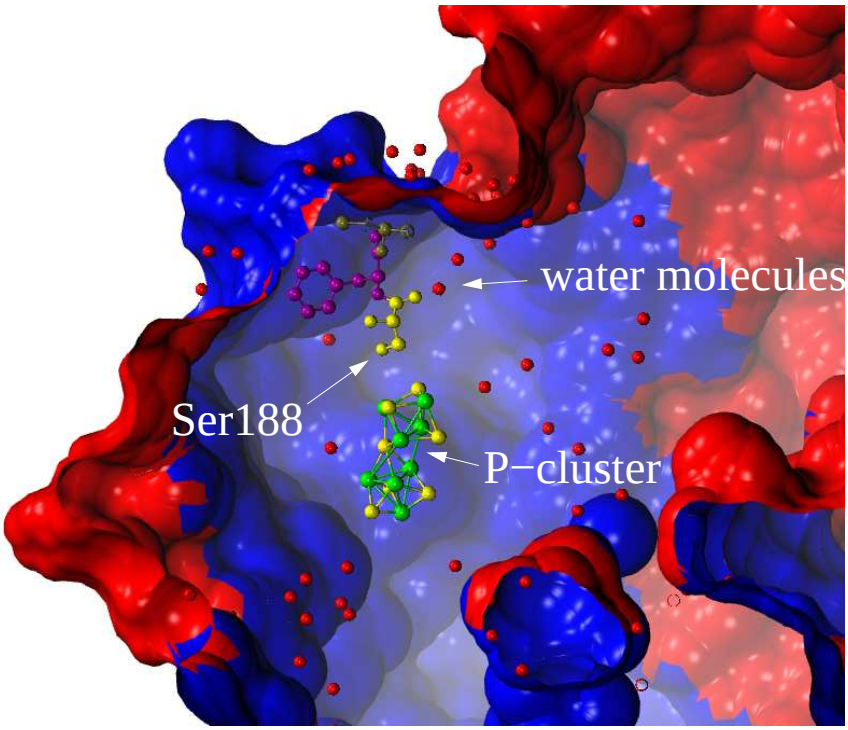


Figure 6.9: Serine-188 residue next to the P-cluster as part of the bridge to the the protein surface. The surface of the protein is depicted by the red and blue area. The different colors label different protein chains. Serine-188 lies near the protein surface. The next water molecule is about 3.5Å away from the oxygen atom of serine. The water molecules are represented by the red spheres.

cluster, the proton pathway offers a new possibility. It has to be investigated whether the structural changes of the P-cluster under change of its oxidation influence the pathway.

6.3. Structural Changes

6.3.1. The resting state P^N

My first calculations of the P-cluster have simulated its resting state. The structure and the total spin of that system can be compared to experimental data. The calculated total spin of the system is S=0 and matches the experimental data perfectly. The resulting structure is compared to the experimental data published in [131]. I will only compare the Fe-Fe distances and the distance of the central sulfur atom to the iron atoms. Table 6.3 shows the Fe-Fe distances of the P-cluster in the 3MIN structure. The distances are in angstrom. FE1-FE8 are the labels used in [131]. FE11-FE14 and FE19-FE23 are the corresponding labels in my calculations.

The table also shows the Fe-Fe distances in the calculation of the resting state. The deviation from the 3MIN structure is given in the second line. The distances

PDB-label	amino Acid
N_{H1}	Arg
N_{H2}	Arg
$N_{\delta 2}$	Asn
$O_{\delta 1}$	Asn
S_{γ}	Cys
$N_{\epsilon 2}$	Gln
N_{Z}	Lys
O_{γ}	Ser
$O_{\gamma 1}$	Thr
O_{H}	Tyr

Table 6.1: Amino acids offering mobile proton transport groups and the PDB-label of these atoms.

from	to	distance in Å
CYS88:N	HOH:108	3.79
HOH:108	HOH:14	2.90
HOH:14	ASN113:ND2	2.78
ASN113:ND2	ASN113:OD1	2.24
ASN113:OD1	HOH:215	2.67
HOH:215	HOH:116	2.76
HOH:116	HOH:2500	2.73
HOH:2500	HOH:257	2.41
HOH:257	HOH:597	2.90
HOH:597	HOH:586	2.79
HOH:586	ASN98:OD1	2.85
ASN98:OD1	ARG96:NH1	3.15
ARG96:NH1	ARG96:NH2	
ARG96:NH2	S5A	3.28

Table 6.2: Proposed proton pathway from P-cluster to FeMo-cofactor.

deviate from the experimental data up to 13 percent. The largest deviations are in distances between iron atoms which belong to different cubanes. The reason for these deviations is that one cubane tilted away. All iron-iron distances within the cubanes agree with experimental data with a deviation of 1-5%. This is a satisfactory result. It has to be remarked that the carbon atoms of the cysteine residues have been fixed in this calculation. The calculation yields a total spin of S=0. Figure 6.11 shows the spin structure of the P-cluster in the resting state.

Although the total spin in the first calculation agrees with the experiment, releasing the carbon atoms of the cysteine residues lead to a non-zero total spin (S=0.3).

	FE 1 FE11	FE 2 FE12	FE 3 FE13	FE 4 FE14	FE 5 FE19	FE 6 FE20	FE 7 FE21	FE 8 FE22	S_1 S_15
FE11		2.662	2.769	2.619	3.665	4.499	5.188	2.520	2.306
		-0.092	-0.071	-0.099	0.047	-0.028	0.132	0.330	0.082
FE12			2.608	2.660	4.730	5.243	6.859	4.514	2.666
			0.072	-0.094	-0.284	-0.521	-0.324	-0.036	-0.245
FE13				2.624	5.174	6.622	7.365	5.111	4.110
				0.051	0.126	-0.076	0.182	0.285	0.035
FE14					2.802	4.487	5.339	3.760	2.399
					0.014	-0.017	0.029	0.108	0.007
FE19						2.490	2.683	2.578	2.493
						0.077	0.012	-0.058	-0.115
FE20							2.652	2.685	2.635
							0.017	-0.126	-0.204
FE21								2.878	4.286
								-0.206	-0.144
FE22									2.328
									0.082

Table 6.3: Distances of the iron atoms of the P-cluster in the 3MIN structure. In the second line the deviations of the distances in my calculated structure to the distances in the PDB are shown. The first row in the head of the table contains the corresponding labels of the atoms in the PDB file. The labels used in my calculations are in the second row.

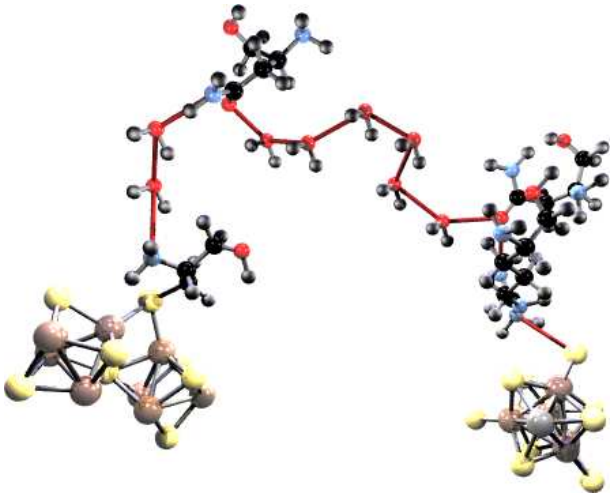


Figure 6.10: Proton path from CYS88:N to FeMo-cofactor. The left iron-sulfur cluster is the P-cluster, the FeMo-cofactor is the cluster on the right side.

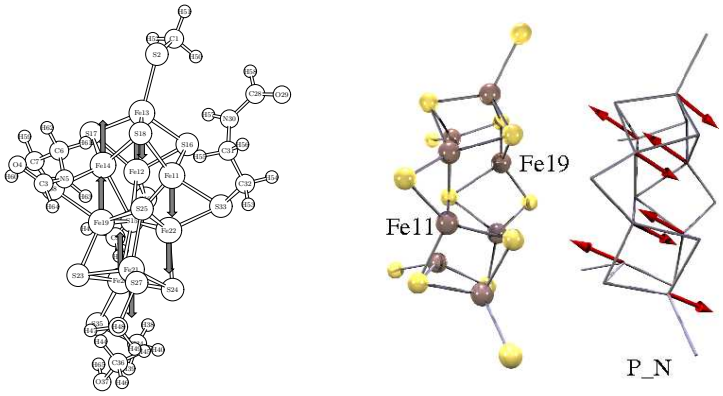


Figure 6.11: Calculated spin structure of P^N from 3MIN structure.

6.3.2. Symmetries in the spin structure

Although the P-cluster has been simulated in a non-collinear calculation with a total spin S=0, we are not sure that the final spin structure is the optimal one. Using symmetry arguments I have constructed a test set of P-clusters with different spin distributions which all have a total spin S=0. These structures are drafted in Figure 6.12. They are labeled as T1 to T7.

In the following calculations the P-cluster is simulated without the surrounding protein environment. Only the sulfur atoms of cysteine residues and the Fe_8S_7 cluster are part of the system. The cysteine sulfur atoms are saturated by hydrogen atoms. Figure 6.13 shows the model of the P-cluster with the labels of the atoms. The hydrogen atoms are not depicted in the figure. The labels of the atoms are the same as in the calculations. I have used these model systems for their simplicity. Trying to understand which influence the spin distribution has, I focused the calculations only on the bare cluster.

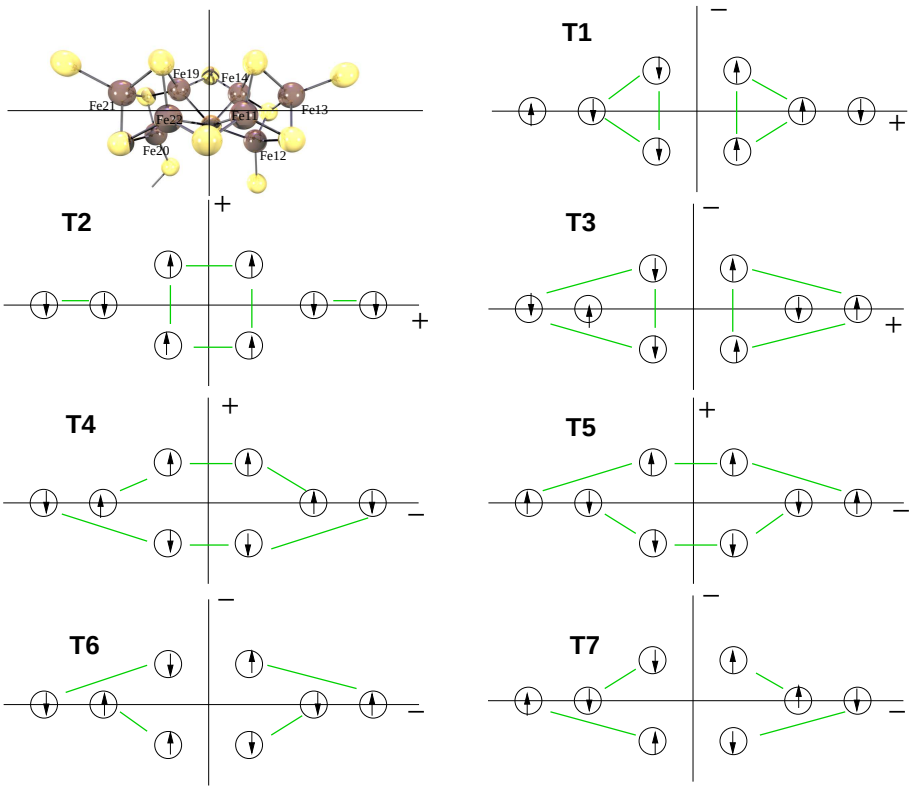


Figure 6.12: All possible spin combinations for $S = 0$ in the P-cluster. + stands for a symmetric, - for an antisymmetric operation at the mirror plane. Green lines between two atoms symbolize a ferromagnetic coupling.

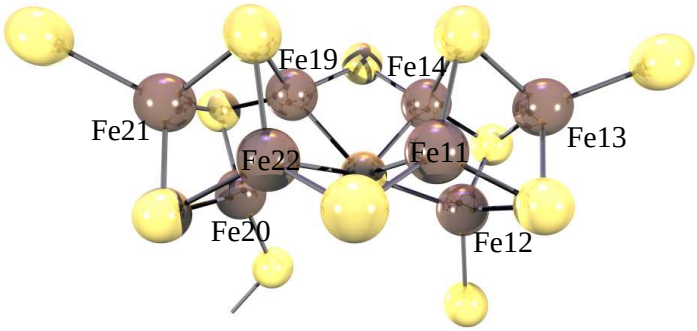


Figure 6.13: The P-cluster without the environmental atoms. Only the sulfur atoms of the cysteine residues are added to the simulation cell. The additional sulfur atoms are saturated by hydrogen which are not shown in the picture. The labels of the atoms are the same as the labels in the calculations.

		FE atom							
Name	Energy [kJ/mol]	11	12	13	14	19	20	21	22
T1	0.00	u	u	d	u	d	d	u	d
T2	60.62	u	d	d	u	u	d	d	u
T3	62.28	u	d	u	u	d	u	d	d
T4	66.77	u	d	u	d	d	d	u	u
T5	67.37	d	d	u	u	u	d	u	d
T6	83.31	d	d	u	u	d	u	d	u
T7	712.51	d	u	d	u	d	d	u	u

Table 6.4: Spin structure and energies of the different spin models of the P-cluster with S=0. The structure of T7 is not fully relaxed only the electrons have been converged. See text for details.

T4 has the same spin structure as the result from the non-collinear calculation of P^N. All systems (T1-T7) are simulated in a collinear calculation in which the distribution of the spins is defined by the user[2]. The table which shows the differences of the Fe positions to the original P^N structure can be found in Appendix E. T7 cannot be found in the table, for the calculation crashes each time the atoms are released. I have not been able to find the reason for these crashes. Comparing the results with Table 6.3 shows that T1 lies closest to the original structure. Table 6.4 shows the energies of all spin models. It can be seen that the more the symmetry in the coupling of the iron atoms is disturbed, the more the energy of the system becomes higher.

T1 has the lowest energy although it does not have the spin structure which

[2] In the CP-PAW program package the !ORBPOT command can be used in the structure file to pre-define the spin.

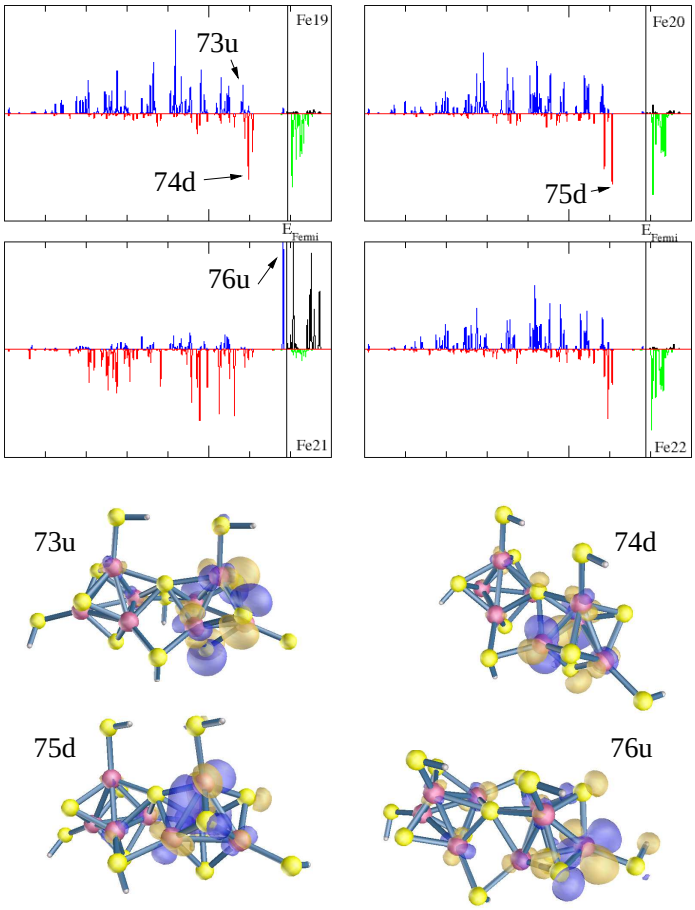


Figure 6.14: The density of states of the iron atoms in the second cubane of the P-cluster in the T1 structure. Only the d-states are plotted. A small set of d-states has been chosen and plotted as wavefunction.

we calculate in a non-collinear calculation.

The density of states of the d-electrons of the iron atoms in the second cubane (FE19 - FE20) is shown in Figure 6.14. Also some of the wavefunctions are plotted which show the iron-iron bond inside the cluster. The iron atoms 13 and 21 show a strong peak of an occupied orbital near the Fermi level. In Figure 6.14 the state 76u is one of them. The wavefunction shows a d-type orbital which has no bonding overlap with the neighboring orbitals.

The cage opens

Before I report the results from the simulation of the T1 spin distribution applied to the P^N system, I will take a closer look at the T3 structure. T3 lies 62.82 kJ/mol higher in energy than T1 but is still the third in the energy scale of the T-systems.

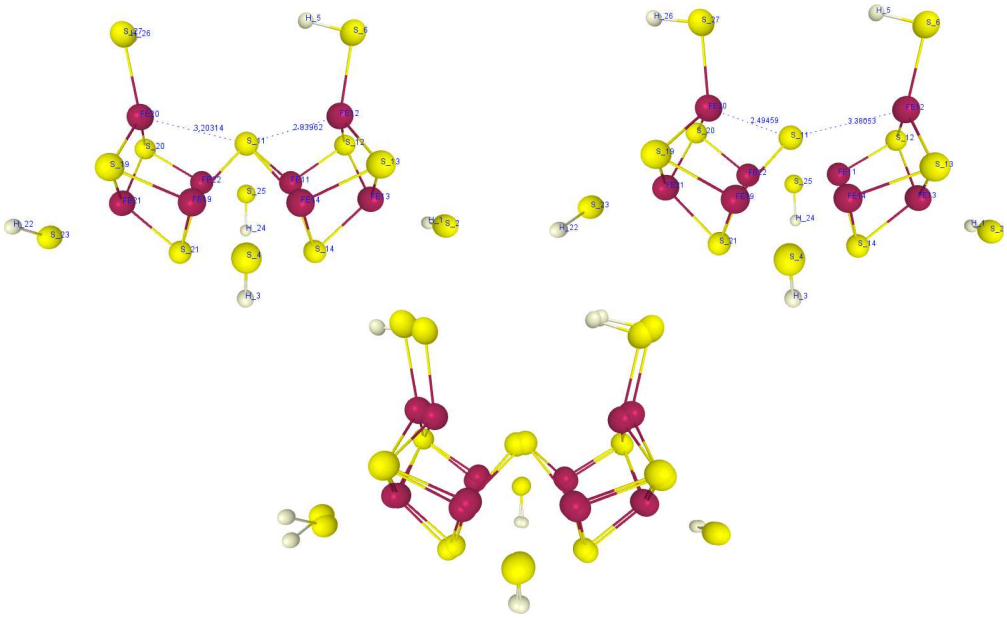


Figure 6.15: T3 in the original and in the flipped form. The third picture shows them both together.

T1 and T3 have in common that, although the total spin of the cluster is S=0, the spin of each cubane is S≠0. All other T-structures have S=0 in each cubane. Interesting about the T3 structure is, that the cluster breaks its symmetry. The bond between the central sulfur atom and Fe20 breaks. This leads to a structure quite similar to the P^{ox} structure. A simulation with a constraint, which forces Fe20 back to its original place, breaks the cluster at the opposite site. The spin structure is the same in both systems. The cluster loses its symmetry in this spin distribution. I would have expected that the bonds of both iron atoms to the central sulfur atom break, thus conserving the symmetry of the cluster.

The iron atoms Fe12 and Fe20 are the only iron atoms which are anti-ferromagnetically coupled to all other iron atoms in the same cubane. This weakens the bond between them. Figure 6.15 shows the broken clusters.

To understand this opening of the cage I investigated the correlation between the spin coupling and the distance of two iron atoms. Table 6.5 shows the distances of the iron atoms in all calculated P-cluster models. The mean value of the anti-ferromagnetic iron-iron bonds is 2.7125Å the mean value of the ferromagnetic ones is 2.688Å. These results are located within the uncertainty of the calculation. There is no significant correlation between the coupling of the iron atoms in the P-cluster and the distance between them.

Switching the spin structure in only one cubane leads to a total spin of S=8. In this case the cluster comes back to its original symmetry, the iron sulfur bond does not break. The total energy is 55.48 kJ/mol higher than the energy of T3 with S=0.

The reason for the break in symmetry can be found by looking at the spin distribution in one cubane. In both structures, T1 and T3, three iron atoms are ferromagnetically coupled and one iron atom is anti-ferromagnetic to all other

	Spin Structure						
Fe-Fe pair	T1	T2	T3	T4	T5	T6	T7
11-12	2.592	2.760	2.772	2.760	2.828	2.692	-
11-13	2.737	2.720	2.616	2.768	2.699	2.699	-
11-14	2.544	2.571	2.487	2.664	2.659	2.624	-
11-22	2.963	2.761	2.983	2.659	2.445	2.746	-
12-13	2.726	2.657	2.750	2.692	2.693	2.698	-
12-14	2.582	2.747	2.770	2.815	2.763	2.695	-
13-14	2.735	2.720	2.613	2.703	2.768	2.789	-
14-19	2.949	2.762	2.976	2.444	2.667	2.709	-
19-20	2.578	2.741	3.482	2.820	2.758	2.717	-
19-21	2.725	2.724	2.581	2.702	2.771	2.739	-
19-22	2.539	2.586	2.470	2.656	2.664	2.614	-
20-21	2.711	2.669	2.672	2.688	2.692	2.651	-
20-22	2.561	2.721	3.470	2.761	2.808	2.584	-
21-22	2.733	2.715	2.591	2.763	2.704	2.702	-

Table 6.5: Distances between iron-iron pairs. Red color labels an anti-ferromagnetic coupling, green color stands for ferromagnetic coupling.

iron atoms of the cubane.

The difference in both structures is that in one case (T1) the iron atom at the edge of the cluster is anti-ferromagnetic coupled to the rest of the iron atoms in the cubane. Looking at the T3 structure in Figure 6.15 one can see that there are three iron atoms in each cubane which lie in a plane while one iron atom is placed above that plane. This iron atom is coupled anti-ferromagnetically to the other iron atoms in the cubane While the three ferromagnetically coupled iron atoms form metal-metal bonds the bond to the anti-ferromagnetic iron atom is weakened. This leads to the break of the bond between the iron atom and the central sulfur atom. A calculation has been done where all distances between iron atoms belonging to the same cubane have been fixed. The cluster stays in its symmetry but the energy is 43kJmol^{-1} higher than the energy of the cluster with broken symmetry. This example shows how the spin distribution is responsible for the geometry of an iron-sulfur cluster. Applying a special spin distribution leads to a break in the symmetry.

The role of the protein environment

As already discussed structure T1 of the test systems lies lowest in energy. In the next step the P-cluster has been calculated with the surrounding atoms of the environment in this spin structure. Again, the carbon atoms of the cysteine residues have been fixed in space. The collinear calculation leads to a total energy of -459.66422H which is 51.85 kJ/mol higher in energy than the energy of P^N. Comparing the Fe-Fe distances to the original structure[3] shows some

[3]Remember: the geometry of T1 has been closest to the original structure.

Structure	NSPIN	free	Total Energy in H	Total Spin in $\hbar$
T4	2	F	-459.58042	0
T4	3	F	-459.58521	0.93780
T1	2	F	-459.61372	0
T1	3	F	-459.61564	0.48865

Table 6.6: The P-cluster in spin structure T1 and T4. The surrounding carbon atoms have been fixed in space. NSPIN=2 stands for a collinear calculation, NSPIN=3 for non-collinear ones.

Total Spin / $\hbar$	Total energy / H	Eigenvalues					
		Spin up			Spin down		
		75	76	77	75	76	77
S=0.0	-314.58423	5.965	6.810	6.902	6.062	6.565	6.594
S=1.0	-314.57944	5.710	6.474	6.557	5.779	6.228	6.383
S=-1.0	-314.57891	5.805	6.196	6.340	5.672	6.500	6.553

Table 6.7: Different spin structures of P-cluster based on T1. Red eigenvalues are occupied, black eigenvalues are unoccupied.

deviations. Calculating the same system in a non-collinear simulation the total spin changes to S=0.3384 and the energy goes down to -459.66628H. This energy is 46.45kJmol^{-1} higher than the energy in the original P^N calculation. In these calculations all atoms have been free. The system has been able to find its optimal geometry. I have made the same calculations with fixed carbon atoms, which forces the P-cluster to keep its original structure. The results can be found in Table 6.6. It can be seen that T1 lies deeper in energy when the environment is fixed in space. However, the total spin in the non-collinear calculations is not zero anymore.

The difference of the energy between the collinear and non-collinear calculations can be explained by looking at the occupied and unoccupied states. At the end of the collinear calculation the lowest unoccupied state in the spin down directions lies deeper in energy than the highest occupied state in spin up direction. Switching to a non-collinear calculation the electron comes down to the deeper lying state which results in a different total spin and a lower total energy. I have made two additional calculations of the P-cluster starting from T1. The total spin has been set to S=1 and S=-1. Table 6.7 shows the results. There is a difference in energy of 1.4kJmol^{-1} between the spin up and the spin down system. The S=0 system has the lowest energy.

6.3.3. The way to the oxidized state P^{ox}

Examining the structural changes of the P-cluster is based on the knowledge from experimental data. The following facts are known:

- P^{ox} has two electrons less than P^N
- the total spin changes from S=0 to S=3 or S=4

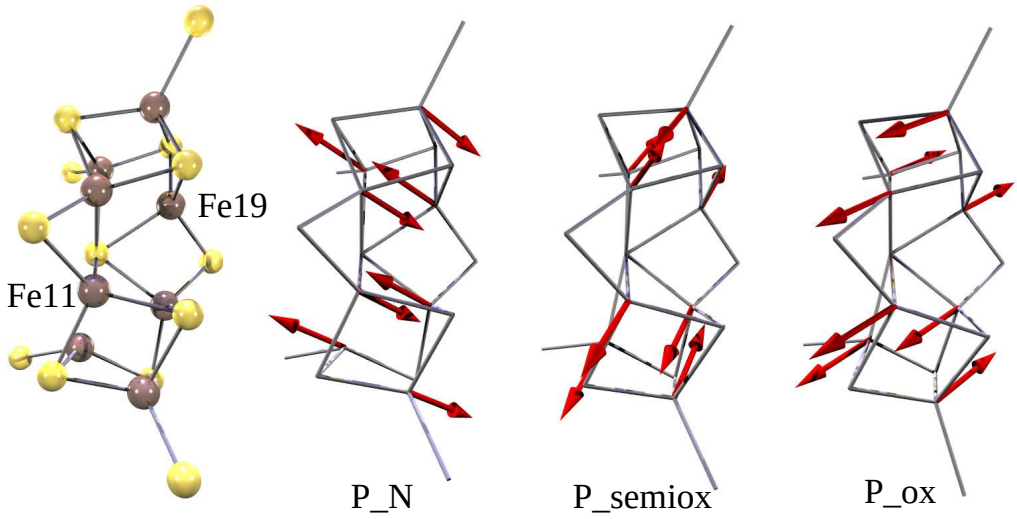


Figure 6.16: Spin distribution of the three oxidation states of the p-cluster

- the cluster builds two new bonds to the protein environment
  - the nitrogen atom of cysteine-88 builds a bond to Fe19
  - the oxygen atom of serine-188 builds a bond to Fe20

Starting from the 3MIN PDB structure the P-cluster has been simulated in its resting state (P^N), in an oxidized state ($P^{semi-ox}$) and in the twofold oxidized state (P^{ox}). Table 6.8 shows the results of the simulations. Figure 6.16 shows the spin structure of the three systems.

In Section 6.2.3 the possibility of a proton traveling from the cysteine residue to the FeMo-cofactor has already been discussed. Also the possibility of protons getting from the solvent to the serine residue. Forming two additional bonds with the environment requires each residue to give away one proton. Based on these facts several calculations have been done which started from the optimized P^N structure:

1. without proton 63 (cysteine residue) and one electron removed
2. without proton 65 (serine residue) and one electron removed
3. without protons 63 and 65 and two electrons removed

System	Total Energy / Hartree	Total Spin / $\hbar/2$
P^N	-459.68397	$S = 0$
P^{semiox}	-459.87277	$S = 0.069$
P^{ox}	-459.93002	$S = 3$

Table 6.8: Total energy and total spin of the three different oxidation states of the P-cluster.

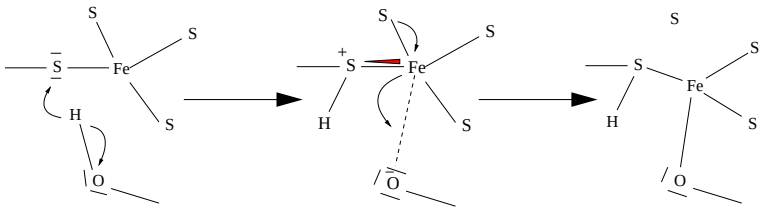


Figure 6.17: Mechanism of the protonation of the sulfur atom 35 in the P-cluster. After protonation FE20 builds a bond to the oxygen of the serine residue.

System	Total Energy / Hartree	Total Spin / $\hbar$
P_N 63	-459.26825	$S = 3.0$
P_N 65	-459.27735	$S = 4.0$
P_N 63+65	-458.74274	$S = 3.0$
P_ox S_35	-459.31757	$S = 3.0$

Table 6.9: Total energy and total spin of the P-cluster with different numbers of electrons and protons.

In the first case the nitrogen atom of the cysteine residue forms a bond to the FE11 atom of the cluster. The iron atom breaks its bond to the central sulfur atom but does not come back to a tetrahedral configuration. The three surrounding sulfur atoms lie in a plane together with the iron atom while the nitrogen atom lies above this plane. In the second case the oxygen atom of the serine residue forms a bond to FE20. The iron atoms breaks its bond to the central sulfur atom and stays in a tetrahedral configuration. In the third case the results are brought together. But the serine oxygen atom formed a bond with a sulfur atom from the nearby cysteine residue. Using constraints in the simulation the oxygen atom has been directed to FE20. The final structure is the well known structure of the P^{ox} state from the 2MIN PDB structure.

A final calculation has been done starting from the P^{ox} state. Again the cysteine proton is removed from the system. Sulfur atom 35 of the P-cluster is protonated by the proton from the serine residue. The remaining serine forms a bond with FE20. The iron atom breaks its bond to the central sulfur atom and the system remains in the correct structure. The total spin of the final structure is S=3 which agrees with experimental data. The mechanism of this reaction is shown in Figure 6.17. The bond between FE20 and the protonated sulfur atom is polarized. The sulfur atom drains electrons from the iron atom which attracts the oxygen from the serine residue. After forming a bond with the oxygen atom, the iron atom breaks its weakest bond which is the bond to the central sulfur atom. Figure 6.18 shows the spin structures of the systems. Table 6.9 shows the energies and the total spin of the discussed systems. The spins of the iron atoms in the final structure are distributed in the same way as they have been reported by Mouesca et. al. [135][4].

[4] Although Mouesca and coworkers used a Fe_8S_8 model for the P-cluster.

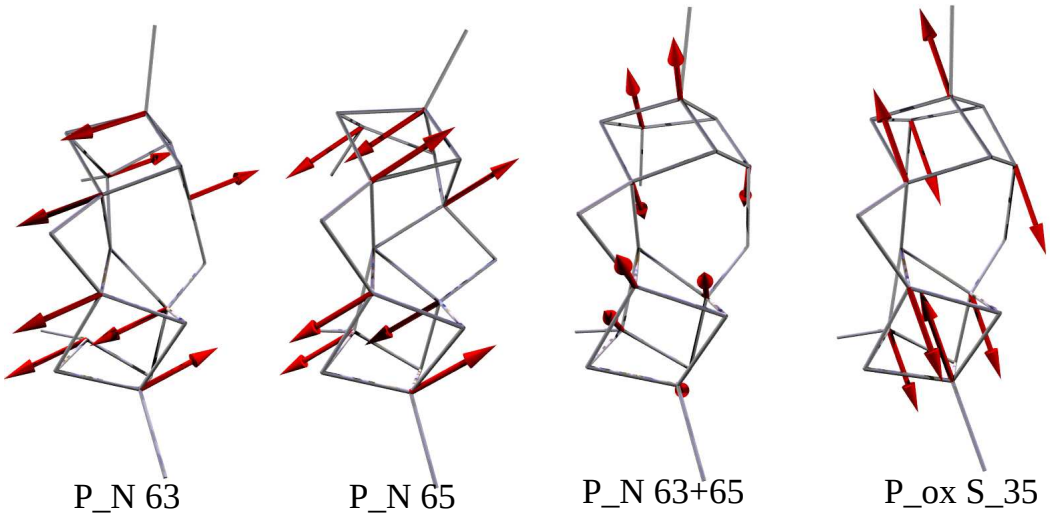


Figure 6.18: Spin structures of the P-cluster with different numbers of protons and electrons. The basic structure is the same as in Figure 6.16

.

6.3.4. The reduced state of the P-cluster

Understanding of the structural changes of the P-cluster after losing two electrons did not help to understand the electron transfer from the P-cluster to the FeMo-cofactor. Therefore, I have investigated the reduced state of the P-cluster.

Figure 6.7 shows that there are the states $P^N_{H^+}$ and P_{H^+} which are also of interest. The first system I calculated has been the $P^N_{H^+}$ system. The whole P-cluster together with atoms of the protein environment as it is described in Section 6.2.2 is part of the simulation cell. The calculations have been non-collinear.

What I have learned from the investigation of the symmetries in the spin structure of the P-cluster is that the iron atoms at the edges of the cluster (FE13 and FE21) have occupied d-orbitals which have an anti-bonding character to all surrounding atoms. Placing a hydrogen atom (proton plus electron) in the vicinity of FE21 yields an interesting structure. The hydrogen atom is placed above the FE20-S25 bond which is shown in Figure 6.19.

A question of interest is why the hydrogen atom is placed above the bond. One possibility would be that the hydrogen is placed between the iron and sulfur atom, thus forming a three-center bond. Figure 6.20 shows a bonding and an anti-bonding three-center bond between an iron d-orbital, the hydrogen s-orbital and a sulfur p-orbital. The calculation yields wave functions which can be interpreted as a mixture of the two three-center bonds in Figure 6.20. The resulting wave functions are shown in Figure 6.21. Both states are occupied. The total spin of the system is S=0.5.

In the next step an electron has been removed from the system. It is called P_{H^+}. The total spin changes back to S=0. Table 6.10 shows the spin distribution of the iron atoms in the P^N, $P^N_{H^+}$ and P_{H^+} structure. Although the total spin

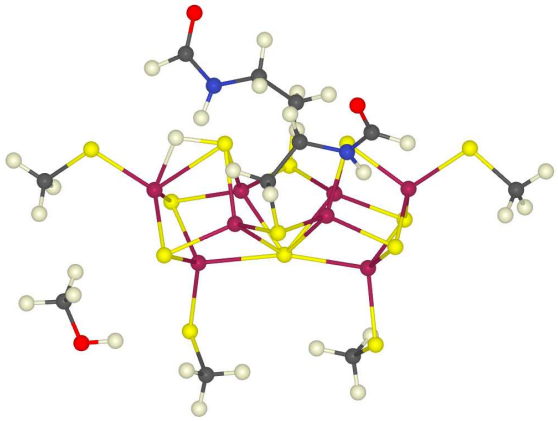


Figure 6.19: The $P^N_{H^+}$ structure of the P-cluster. The hydrogen atom is located above the FE20-S25 bond.

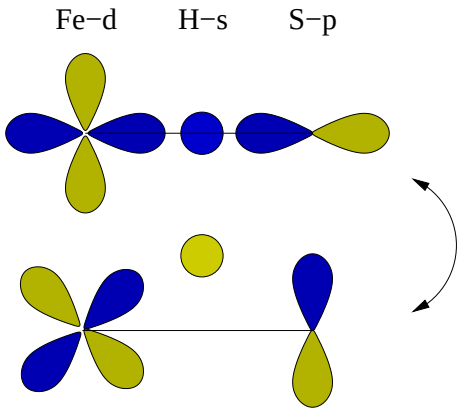


Figure 6.20: A bonding and an anti-bonding three-center bond between an iron d-orbital, a hydrogen s-orbital and a sulfur p-orbital.

of the cluster is zero in the P^N and the P_{H^+} state, the internal spin structure changes. In the P^N state each cubane has a total spin S=0. Figure 6.22 shows a symmetry in the coupling between the iron atoms with respect to the vertical axis. Ferromagnetic coupling is labeled with green bars and anti-ferromagnetic coupling is depicted by red bars between the iron atoms. Regarding the horizontal axis the coupling is nearly anti-symmetric. The symmetry is broken, for the coupling between Fe11 and Fe22 is ferromagnetic. Adding one proton to the cluster, the spin structure changes. In the new arrangement, the total spin remains zero, while each cubane has a non-zero spin. The new spin structure is anti-symmetric with respect to the vertical axis, while the coupling is symmetric. Regarding the horizontal axis the spin structure and the coupling between the iron atoms is symmetric. In Section 6.3.2 different spin structures of the P-cluster have been discussed. Model T1 is the spin structure which is energetically most favorable. Its spin structure is identical with the spin distribution of the P_{H^+} state.

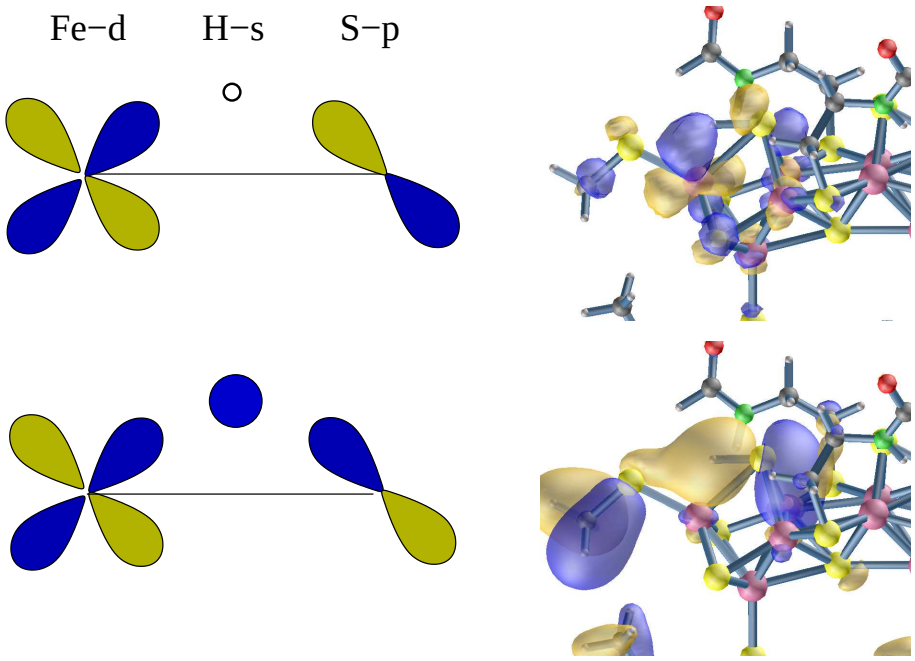


Figure 6.21: The bonding and the anti-bonding orbital of the Fe-H-S bond. Both orbitals are occupied.

System	first cubane				second cubane				total spin
	11	12	13	14	19	20	21	22	
P^N	↓	↑	↓	↑	↑	↑	↓	↓	S=0
$P^N_{H^+}$	↑	↑	↓	↑	↓	↓	↑	↓	S=0.5
P_{H^+}	↑	↑	↓	↑	↓	↓	↑	↓	S=0

Table 6.10: Spin distribution of the iron atoms in different protonation and oxidation states of the P-cluster.

6.4. Conclusions and Outlook

In this Chapter I have presented a closer look at the P-cluster and its spin distributions. I have examined different spin structures in the resting state as well as the spin distributions in different oxidation states. Comparing these results with our knowledge of the mechanism of the nitrogen fixation at the FeMo-cofactor, I can draw a new picture of the role of the P-cluster. Electrons and protons are needed during the catalytic cycle. It is assumed that they are transferred in a correlated way. In my picture they both are transferred from the P-cluster to the FeMo-cofactor. Experiments give no information about protons at the P-cluster in the oxidized state. From the model systems I have learned that there are two possible iron atoms where a proton can be placed. One of these proton sites has been simulated and analyzed by looking at the P_{H^+} system. There, a proton is placed between Fe21 and S25. This state has the same spin distribution as the P^{ox} state. I propose that the P-cluster has two additional protons at Fe13 and Fe21 in its oxidized state. These protons

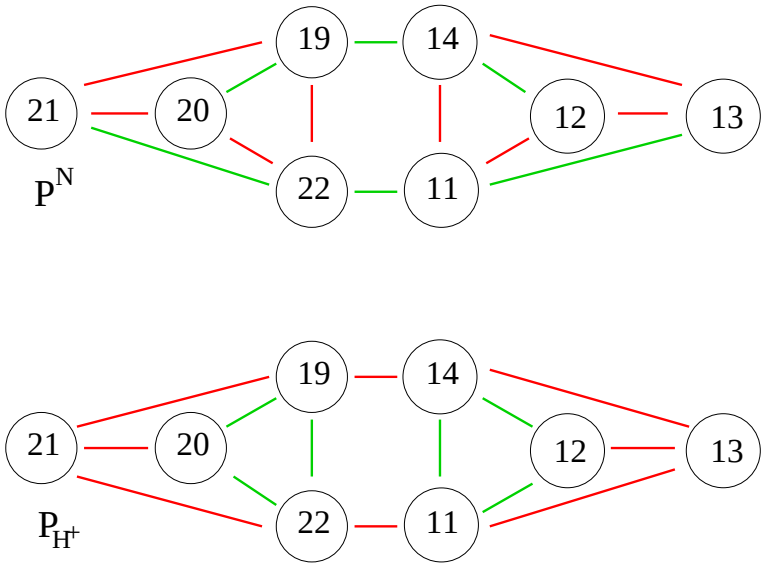


Figure 6.22: The coupling between the iron atoms of the P-cluster in the resting state P^N and with one proton added, P_{H^+}. Green lines label ferromagnetic coupling between the iron atoms, red lines symbolize anti-ferromagnetic coupling.

can be transported to the FeMo-cofactor as needed.

The next step should be to simulate the cluster with the protein environment in a QM/MM simulation. In a following step the program code should be extended so that a QM/MM simulation with two QM centers could be performed. The parallelization technique presented by Poddey [140] can be used to realize the two QM centers.

7. The Continuum Solvation Model - COSMO

Besides our work on implementing the Cornell et. al. force field into the CP-PAW program package to simulate structures including the solvent, we started a project together with BASF to include another approach for simulating solvents. The goal of this project is the implementation of the conductor like screening model (COSMO) into the current version of the CP-PAW program package and its extension to periodic systems. COSMO has been developed by Klamt and Schüürmann [141]. Senn et. al. reported an implementation within the framework of Car-Parrinello *ab-initio* molecular dynamics [142]. In order to permit a seamless integration of non-periodic description with the extension to periodic systems, we decided in favor of a complete rewrite of the COSMO implementation. The existing COSMO implementation has been incorporated as well, in order to allow for rigorous testing of the new approach.

7.1. The energy expression

COSMO is a method for the description of solvent effects in first-principles calculations. This method embeds a solute into a continuum dielectric that describes the screening by a polarizable solvent. The dielectric is described by surface charges that are dispersed about the surface of the solute and that interact electrostatically with the solute. In addition a correction term for non-polar interactions is included.

The COSMO solvation energy has the form:

$$
\begin{aligned}
E &= \sum_i q_i \Theta_i \int \mathrm{d}^3 r \, \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{s}_i|} && (7.1.1)\\
&+ \frac{1}{2f} \sum_{i \neq j} \frac{q_i \Theta_i q_j \Theta_j}{|\mathbf{s}_i - \mathbf{s}_j|} && (7.1.2)\\
&+ \frac{1.07\sqrt{\pi}}{f} \sum_i \frac{q_i^2 \Theta_i^2}{\sqrt{a_i}} && (7.1.3)\\
&+ k \sum_i (1-\Theta_i) q_i^2 && (7.1.4)\\
&+ \beta + \gamma \sum_i a_i \Theta_i && (7.1.5)
\end{aligned}
$$

The screening charges are denoted by q_i and their positions by $\mathbf{s}_i$. The dielectric

constant is described by the factor

$$f = \frac{\varepsilon_r - 1}{\varepsilon_r} \tag{7.1.6}$$

a_i is the area of a triangular face related to a surface charge. $\rho(\mathbf{r})$ is the solute charge density which is built by atom-centered Gaussians.

The term in (7.1.1) describes the direct solvent-solute interaction. (7.1.2) describes the electrostatic interaction between different surface charges. The surface charges are point charges. Due to the $1/r$ behavior of the Coulomb interaction, these point charges can lead to large destabilizing energies, if the charges come too close. Therefore, the Coulomb interaction is replaced by a linear function:

$$\frac{1}{r} \longrightarrow \begin{cases} \frac{2}{R^c_{i,j}} - \frac{r}{(R^c_{i,j})^2} & \text{for} \quad r < R^c_{i,j} \\ \frac{1}{r} & \text{for} \quad r > R^c_{i,j} \end{cases} \tag{7.1.7}$$

The cut-off radius $R^c_{i,j}$ is determined by

$$R^c_{i,j} = \frac{2f}{1.07\sqrt{\pi}}(a_i^{-1/2} + a_j^{-1/2})^{-1} \tag{7.1.8}$$

as it is described in [142]. The electrostatic self energy of the surface charges is given by (7.1.3).

Charges lying in the overlap region of another atom do not contribute to the COSMO energy. The term in (7.1.4) is used to determine whether the charge contributes to the COSMO energy or not. The switching factors Θ_i are defined as

$$\Theta_i = \prod_{J \neq I(i)} \eta\left(|\mathbf{R}_J - \mathbf{s}_i| - R^{solv}_J\right) \tag{7.1.9}$$

$$\eta(x) = \left(1 - \mathrm{e}^{-(x+c)^{2n}}\right)\theta(x+c) \tag{7.1.10}$$

with $n = 24$ and $c = 0.9$. R^{solv}_J is the solvation radius of atom J. $I(i)$ denotes the atom I to which the surface charge belongs. Figure 7.1 shows $\eta(x)$ for different parameters n[1].

The term in (7.1.4) describes the nonpolar interaction which is proportional to the solvent accessible surface area. The parameters for the non-polar interactions are $\beta = 2.1567 \cdot 10^{-3}$ and $\gamma = 6.3739 \cdot 10^{-6}$.

For the parts (7.1.1)-(7.1.3) of the COSMO energy expression, the concept of an effective charge $\bar{q}_i = q_i\Theta_i$ shall be introduced in order to expand the

[1] The sign of parameter c in our implementation is opposite of the sign in the literature. The original η-function is defined as

$$\eta(d, d^0) = 1 - \exp[-(d - d^0)^{2n}]$$

d^0 is defined as $d^0 = R^{solv} - c$. The argument of the η-function in our code is DISS-RSOLV which leads to the correct term.

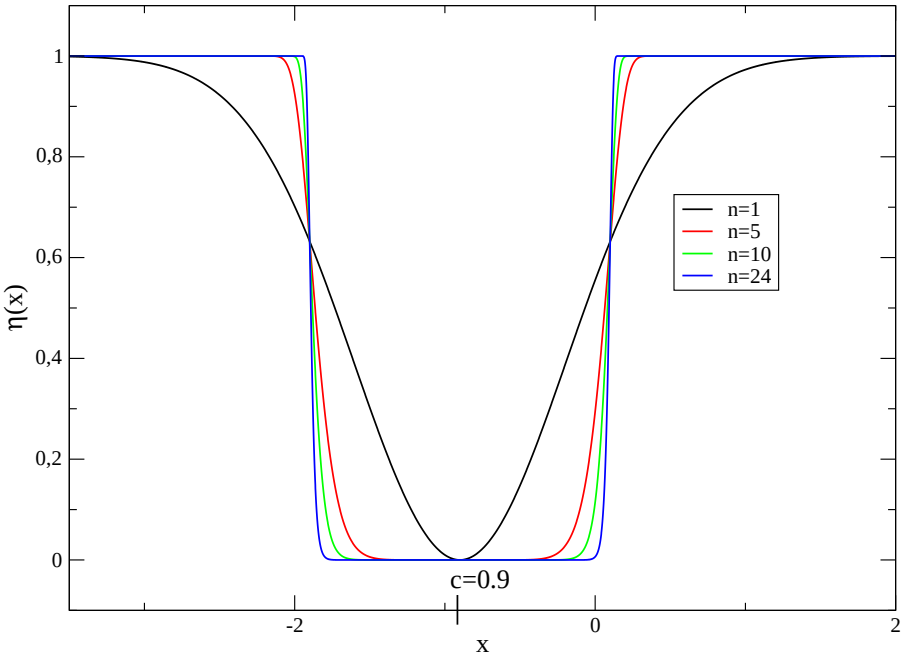


Figure 7.1: η-function to switch charges off which do not contribute to the COSMO energy. $\eta(x)$ is printed for different parameters n. The parameter c is set to 0.9.

implementation to periodic systems.

$$\frac{\partial E}{\partial q_i} = \frac{\partial E}{\partial \bar{q}_i}\Theta_i \tag{7.1.11}$$

$$\frac{\partial E}{\partial R_I} = \left.\frac{\partial E}{\partial R_i}\right|_{\bar{q}} + \sum_i \frac{\partial E}{\partial \bar{q}_i}\frac{\partial \Theta_i}{\partial R_i} \tag{7.1.12}$$

Thus, the energy expression in equation (7.1.1)-(7.1.5) can be rewritten in the following form:

$$\Delta E_c = \frac{1}{2f}\left[\sum_{i\neq j}\frac{q_i\Theta_i q_j\Theta_j}{|\mathbf{s}_i-\mathbf{s}_j|} + 2\sum_{i,j}\frac{q_i\Theta_i fQ_J}{|\mathbf{s}_i-\mathbf{R}_j|} + \sum_{I\neq J}\frac{fQ_i fQ_j}{|\mathbf{R}_I-\mathbf{R}_J|}\right] \tag{7.1.13}$$

$$-\frac{1}{2f}\sum_{I\neq J}\frac{fQ_i fQ_j}{|\mathbf{R}_I-\mathbf{R}_J|} \tag{7.1.14}$$

$$+\frac{1}{f}\sum_i q_i\Theta_i\int \mathrm{d}^3r\,\frac{f\rho(r)-\sum_J fQ_J\delta(r-R_i)}{|\mathbf{r}-\mathbf{s}_i|} \tag{7.1.15}$$

$$+\frac{1.07\sqrt{\pi}}{f}\sum_i\frac{q_i^2\Theta_i^2}{\sqrt{a_i}} \tag{7.1.16}$$

$$+k\sum_i(1-\Theta_i)q_i^2 \tag{7.1.17}$$

$$+\beta+\gamma\sum_i a_i\Theta_i \tag{7.1.18}$$

The atomic positions are denoted by $\mathbf{R}_I$. The charges Q_I are built by atom-centered Gaussians charges. These Gaussians are fitted to the charge density of the I-th atom.

7.2. Construction of the cavity

To simulate the solvent a cavity is constructed. A sphere with a well-defined solvation radius R_I^{solv} is assigned to the I-th atom of the system. The surface of such a sphere is triangulated. That means that the surface is segmented into t_I isosceles triangles of area a_I. The basic form of the triangulations is to build sixty triangles forming a pentakis-dodecahedron structure. The surface charge is placed in the midpoint of each segment. Figure 7.2 shows the sixty surface charges around one hydrogen atom. The tessellation of the sphere can improved by dividing each triangle into four smaller ones. By default CP-PAW uses sixty surface charges. The user can specify the tessellation file if he wants to use a finer grid. Tessellation files with 240, 960 and 3840 surface charges are included in the program package.

If two atoms come close enough and the spheres overlap the Θ functions blanks out the charges lying within another sphere. Figure 7.3 shows a water molecule lying inside a cavity. The positive surface charges are blue, the negative are red.

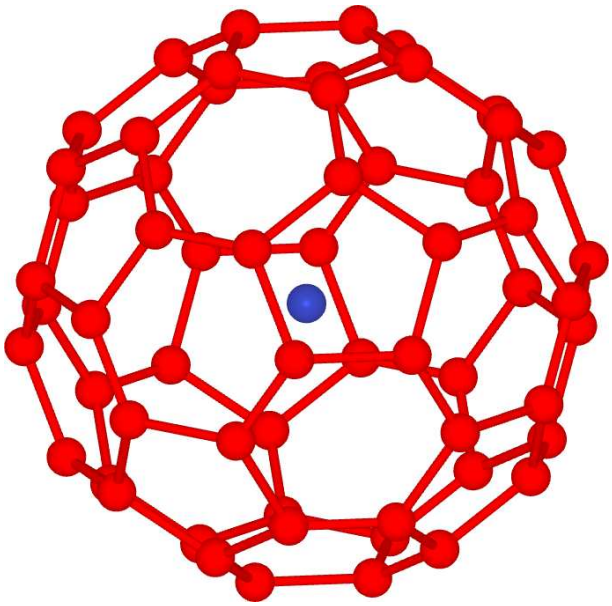


Figure 7.2: A hydrogen atom (blue) surrounded by 60 point charges (red).

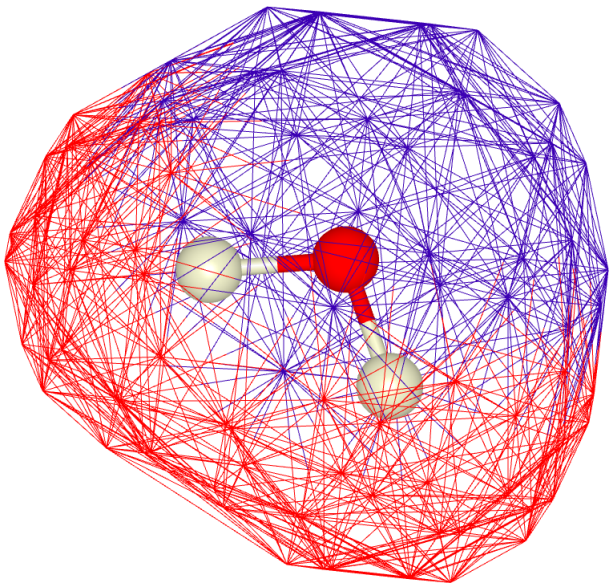


Figure 7.3: A water molecule lying in the cavity. The surface charges are colored blue in the case of positive charges and red in the case of negative charges. The solvation radius has been chosen arbitrarily (3.6 a_0 for the oxygen atom and 2.8 a_0 for the hydrogen atoms).

Variable	Value
NAT	3
POS	$\begin{pmatrix} 1.59013892626677 & 1.1871410827355 & 0.139811988219690 \\ -0.607902197826223 & -1.18080753120953 & -0.0944110980136977 \\ 0.269560771146478 & 0.924920798737901 & -0.155778026320951 \end{pmatrix}$
NG	4
RC	$\begin{pmatrix} 0.1510 & 0.1510 & 0.3450 \\ 0.5000 & 0.5000 & 0.5000 \\ 0.7000 & 0.7000 & 0.7000 \\ 0.9800 & 0.9800 & 0.9800 \end{pmatrix}$
QI	$\begin{pmatrix} -0.996382088883661 & -0.996386004734498 & -5.49549899699248 \\ 5.83665579333337 & 5.37255328421707 & 24.1804484773500 \\ -9.29425933501348 & -8.54319578606125 & -42.9242424586155 \\ 4.01057738634731 & 3.72030951812874 & 25.1294202080908 \end{pmatrix}$

Table 7.1: Input data for the two COSMO implementations passed to CONTINUUM$PROPAGATE (old implementation) and COSMO$INTERFACE (new implementation) in the isolate object.

7.3. Test calculations

7.3.1. Numerical agreement with the previous COSMO implementation

After rewriting the COSMO object, the first step was to test it against the old implementation. We used a single water molecule to test the routines. In paw_isolate one can switch between the new and old COSMO implementations. First we calculated two steps with the previous COSMO implementation, to obtain the input variables for the test from the subroutine CONTINUUM$PROPAGATE. In the next run we pass these values to the COSMO$INTERFACE routine and compare the results for both implementations. Table 7.1 shows the input data and Table 7.2 the output data.

Both implementations use different variable names. To avoid misunderstandings they are listed in Table 7.3.

The results from this test are quite satisfying, there are only minimal deviations. In the next step we test the implementation of the the periodic images in the COSMO framework.

7.3.2. Agreement with experimental data

In the next step we calculated the solvation energies for a set of organic molecules. The results are compared to the experimental data and the results from the former COSMO implementation as they are published in [142]. Each molecule is calculated by standard procedures with the CP-PAW program package without COSMO. The optimized and converged structures are calculated again with COSMO. The difference of the resulting total energies is the solvation energy.

Variable	CONTINUUM$PROPAGATE	COSMO$INTERFACE
VI1 ($\cdot 10^2$)	$\begin{pmatrix} 2.51431 & 2.58064 & -1.70734 \\ 2.51431 & 2.58064 & -1.70734 \\ 2.51425 & 2.58058 & -1.70734 \\ 2.50433 & 2.57055 & -1.70715 \end{pmatrix}$	$\begin{pmatrix} 2.51537 & 2.58199 & -1.70902 \\ 2.51537 & 2.58199 & -1.70902 \\ 2.51531 & 2.58193 & -1.70902 \\ 2.50539 & 2.57189 & -1.70875 \end{pmatrix}$
longrange	-0.026421	-0.026431
EBlank	$5 \cdot 10^{-7}$	$4.64 \cdot 10^{-7}$
ENonpolar	0.0	$2.94814 \cdot 10^{-3}$
ESelf	0.0071389	0.007136057

Table 7.2: Difference of output data after the same data input into CONTINUUM$PROPAGATE (old implementation) and COSMO$INTERFACE (new implementation) in the isolate object.

COSMO variable	HMS variable
longrange	inter triangle
shortrange	ion-triangle
self	intra triangle
blank	cavity hardness
nonpolar	surface tension

Table 7.3: Corresponding variable names in the old (HMS) COSMO implementation and in the new implementation.

System	E_{tot} [H]	E_{tot} COSMO [H]	Diff. [kJ/mol]	HMS	Experiment
H_2O	-17.35486	-17.36524	-27.25	-26.10	-26.40
CH_4	-8.14702	-8.14450	+6.62	+6.60	+8.40
NH_3	-11.84052	-11.84865	-21.35	-20.40	-18.00
MeCOMe	-37.00206	-37.01179	-25.55	-23.40	-16.10
MeOH	-24.28796	-24.29561	-20.90	-18.00	-21.40
PhMe	-45.06515	-45.06432	+2.18	+0.20	-3.70
Pyridine	-41.79096	-41.79870	-20.32	-20.80	-19.70

Table 7.4: First test cases with new COSMO compared to old COSMO as published by Senn [142] (HMS) and experiment.

Table 7.4 shows the test systems and the solvation energies compared to the experimental data and the published results from [142]. The calculated solvation energies of small molecules H_2O, CH_3 and NH_3 in water agree within $2 kJmol^{-1}$ with those published earlier. The small difference can be attributed most likely to the different density functionals used.

7.3.3. The solvation radius

The COSMO energy and thus the free energy of solvation ΔG_s^0 strongly depends on the solvation radius. The solvation radius has to be set for each atom explicitly. As an example the solvation radius of the chlorine atom in 2-chlorobutane and chloromethane has been varied. The resulting energy depends linearly on the solvation radius in a long range. The results are shown in Figure 7.4. In the next step an optimal parameter set of solvation radii is needed. To built this set we set up a data base with small organic molecules. The basic group of molecules in the database are hydrocarbons (to determine the solvation radii for hydrogen and carbon). The other groups add one additional element to the basic group, thus we have one group for oxygen, nitrogen, chlorine and sulfur. These are the main elements in organic chemistry[2]. Table 7.5 shows the test molecules of the first database. Each molecule in the database has been pre-calculated with CP-PAW without COSMO. We wrote a script which automatizes the process. The script produces an output file which contains for each molecule the total energy without COSMO and the energies with COSMO and different solvation radii. We use the vdW radius of the atoms and a variation of the vdW radius. Additionally, the experimental solvation energy has to be included in the result file for each molecule. Finally, a least square fit yields the new optimized solvation radii for each element. Table 7.6 shows the optimized solvation radii for the test set. These solvation radii can be used to repeat the COSMO calculations. However, the new calculated solvation energies are not very accurate. The main problem is that the test set is too small to yield reasonable parameters. The structure of the database and scripts to evaluate the parameters are built in a way to expand the database to more systems easily. We passed this package to BASF. They can use their own database of structural input data together with

[2] Phosphorus is still missing, but can be easily added to the database.

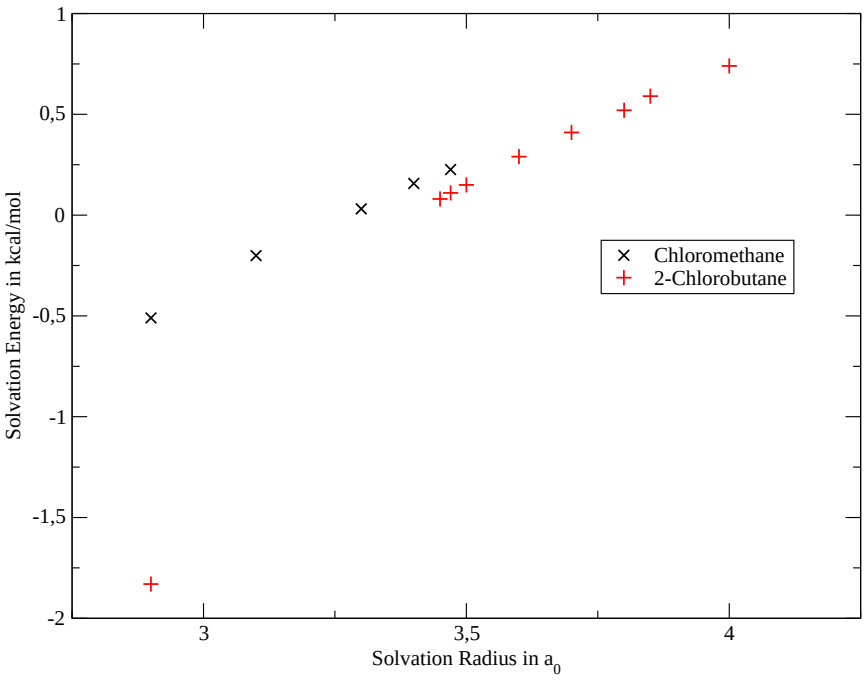


Figure 7.4: The influence of the solvation radius of a chlorine atom on the energy of the system. In this example chloromethane and 2-chlorobutane have been calculated.

Element	Molecule	Element	Molecule
carbon	methane	chlorine	1,1,1-trichloroethane
	cyclohexane		1,1,2-trichloroethane
	phenylmethane		2-chlorobutane
			chlorobenzene
			chloromethane
			o-dichlorobenzene
			p-dichlorobenzene
oxygen	water	sulfur	diethyl-disulfide
	H_2CO		diethyl-sulfide
	phenol		dimethyl-disulfide
	propanon		dimethyl-sulfide
	THF		methane thiol
	naphtaldion		thioanisole
			thiophene
nitrogen	ammonia		
	pyridine		

Table 7.5: List of organic molecules in the database to determine the optimal solvation radii.

Element	Solvation radius in a_0	
	vdW	optimized
carbon	3.851	3.844
hydrogen	2.886	0.951
chlorine	3.947	3.948
nitrogen	3.660	3.658
oxygen	3.500	4.090
sulfur	4.035	2.193

Table 7.6: The van-der-Waals radii and the optimized results from the database for the calculation of solvation energies with COSMO.

	isolated	periodic
without COSMO	-0.2843603	-0.3022635
with COSMO	-0.3390506	-0.3394448
COSMO energy	-0.0547250	-0.0372208
longrange	-0.0640028	-0.0464980
shortrange	-0.0018331	-0.0011886
self	0.0104612	0.0104700
solvation energy	-0.05469	-0.0371810

Table 7.7: Different energy contributions to the COSMO energy in the periodic and in the non-periodic case. All energies are in Hartree.

our program package to generate better solvation radii.

7.3.4. Periodic systems

The hydrogen atom

In order to test the periodic COSMO implementation we performed a calculation of a non-spin polarized, positively charged hydrogen ion $H^{+\frac{1}{2}}$ in a simple cubic unit cell with lattice constant 20 a_0. The solvation radius has been chosen as 2.2 a_0. The contributions to the total COSMO energy and the difference between the periodic and the non-periodic case are listed in Table 7.7. The energy 'longrange' consists of the terms given in equation (7.1.13) and (7.1.14). The energy 'shortrange' consists of equation (7.1.15) and the energy 'self' consists of equations (7.1.16),(7.1.17) and (7.1.18). The self energy 'self' is practically unchanged. The dominant contribution to the difference between the isolated and the periodic system of 17.5 mH stems from the longrange energy.

Consider the energy between the periodic and the isolated case. The energy difference between an isolated ion and a periodic array of ions in a compensating background is 17.9032 mH without COSMO and only 0.3942 mH with COSMO. Thus the interaction energy between periodic images is reduced by a factor 40, when the screening is taken into account. Since we used a dielectric constant $\varepsilon_r = 80.1$, we would expect that the interaction energy of the periodic solvated system is even closer to the solvated isolated system. However, given

	Q	$\sum_i \bar{q}_i$	$\bar{Q} + \sum_i \bar{q}_i$
isolated	-0.49375788	0.49881989	0.005962
periodic	-0.49375788	0.49910139	0.005344
difference	0.00000000	0.00028200	0.000282

Table 7.8: The difference of the screening charges in the periodic and in the non-periodic COSMO implementation,

that we should also consider the contribution of the compensating background the agreement is satisfactory. Table 7.8 lists the screening charges in both implementations. Thus the screening charges are similar in the periodic and in the non-periodic case. We can now estimate the energy difference between the periodic and the non-periodic calculation via a simple correction, which replaces the unit cell by a sphere with equal volume and ignores the overlaps of the resulting spheres.

$$\Delta E \approx -\frac{9}{10}\frac{(\bar{Q} + \sum_i \bar{q}_i)^2}{4\pi\varepsilon r_V} \tag{7.3.1}$$

where

$$V = \frac{4\pi}{3} r_V^3 = 20^3\,\mathrm{a}_0 \qquad \Longrightarrow \qquad r_V = \sqrt[3]{\frac{3}{4\pi}}20\,\mathrm{a}_0 = 12.407\,\mathrm{a}_0$$

For the solvated ion with $\bar{Q} + \sum_i \bar{q}_i = 0.005$ one obtains $2 \cdot 10^{-3}$ mH. For the unsolvated ion with $\bar{Q} = -0.5$ one obtains -18.14 mH. The agreement with the difference of -17.5 mH between the two longrange contributions of less than 4% is fairly satisfactory given the model assumptions.

Whereas the charge $\bar{Q}$ is identical in the periodic and non-periodic cases, the surface charges deviate slightly. Therefore the interaction energy between the center and the screening charges is evaluated:

$$\Delta E = \frac{\bar{Q}\sum_i \bar{q}_i}{4\pi\varepsilon_0 r_{solv}} \tag{7.3.2}$$

for the difference of ΔE between the periodic and the non-periodic case is 0.063 mH, which is negligible compared to the change of the longrange contribution.

The water molecule

We choose water for a periodic system. One water molecule in an elementary cell of $\begin{pmatrix} 0.00 & 4.95 & 4.95 \\ 4.95 & 0.00 & 4.95 \\ 4.95 & 4.95 & 0.00 \end{pmatrix}$ is calculated periodically. The results, which are obtained from the calculation, are presented in Table 7.9. Again, the results are satisfactory.

System	E_{tot} in Hartree	E_{tot}^{cosmo} in Hartree	Difference in kJmol^{-1}
H_2O	-17.35486	-17.36533	27.49
H_2O periodic	-17.35551	-17.36222	17.62

Table 7.9: A water molecule calculated in a periodic and in a non-periodic cell with COSMO.

Part III.

Appendix

A. Computational Details

All DFT calculations in this work are based on the projector augmented wave method (PAW) as it is implemented in the CP-PAW program package. DFT and PAW have been described in Chapter 4. The gradient-corrected PBE [83] functional has been used for exchange and correlation.

For PAW is a method which is based on plane waves, artificial periodic images occur when simulating none-periodic structures. I used the scheme presented by Blöchl [143] to subtract the electrostatic interactions between these periodic images. I made sure that atoms of neighboring cells are at least separated by 6Å. This avoids an overlap of the wavefunctions across the cell boundary. A plane wave cutoff of 30 Rydberg for the auxiliary wave functions and 60 Rydberg for the density has been used.

The shells treated in the frozen core approximation and the sets of projector functions which have been employed are shown in Table A.1.

Atomic structures have been optimized by using Car-Parrinello molecular dynamics. The structures have been converged applying a friction on the wave functions. The convergence has been tested by checking whether the kinetic energy of the wave functions remains zero for at least 200 time steps (=0.05ps).

A non-collinear description of the spin density has been used in my calculations of the P-cluster. This method avoids artificial barriers between different spin configurations occuring in collinear calculations. The resulting spin structure is independent of the random starting conditions.

The input structures for the calculations have been modeled using MSModelling or they are based on PDB files. The latter have been filled with hydrogen atoms (if necessary) by using 'reduce' or the 'paw_tostrc' tool. The tool from the CP-PAW program package has been modified by the author in a way that

Element	core	s	p	d
H	none	1	0	0
C	[He]	2	2	0
N	[He]	2	2	0
O	[He]	2	2	0
P	[Ne]	2	2	0
S	[Ne]	2	2	0
Cl	[Ne]	2	2	0
Fe	[Ar]	2	2	2

Table A.1: Core configuration and number of projector functions with s,p and d-character as they have been used in the simulations during the work on this thesis.

it can create CP-PAW input structures from PDB or MSModelling files. In the case of PDB files it can add the hydrogen atoms or use PDB files which have been preprocessed by 'reduce'.

Different programs have been used to create the pictures shown in the thesis. From the calculations the final structure is given in a xyz-format. This can be used by a program called 'render_xyz' written by the author of this thesis. The program creates a file which contains the data from the calculation and which can be visualized and rendered with 'povray' [144]. Another useful tool is VMD [145] which is the best choice visualizing protein structures. The algorithm developed by Frishman et. al. [146] has been implemented in VMD which enables the visualization of secondary structures in proteins. The third software used to create pictures from atomic structure data is CryMolCAD which is developed by the IT team in our group.

Wave functions and densities have been plotted using OpenDX [147].

B. The P^N structure file

```
!STRUCTURE
   !GENERIC
     LUNIT=1.8897259926
   !END
   !OCCUPATIONS
     NSPIN=3 CHARGE[E]=-4.0 empty=15
   !END

   !LATTICE
      T=     0.00000   15.00000   15.00000
            15.00000    0.00000   15.00000
            15.00000   15.00000    0.00000
   !END

   !SPECIES
       NAME='H_' ZV=1. M=2. npro=1 lrhox=1
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/001H/h_.75
           _6.0.out'
   !END
   !SPECIES
       NAME='C_' ZV=4. M=5. npro=2 2 lrhox=2
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/006C/c_.75
           _6.0.out'
   !END
   !SPECIES
       NAME='N_' ZV=5. M=5. npro=2 2 lrhox=2
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/007N/n_.75
           _6.0.out'
   !END
   !SPECIES
       NAME='O_' ZV=6. M=5. npro=2 2 lrhox=2
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/008O/o_.75
           _6.0.out'
   !END
   !SPECIES
       NAME='S_' ZV=6. M=5. npro=2 2 lrhox=2
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/016S/s_.75
           _6.0.out'
   !END
   !SPECIES
       NAME='FE' ZV=8. M=5. npro=2 2 2 lrhox=2
       FILE='/fastfs/work/nicsh667/PAW/Setups_031213/PBE/026FE/fe_
           .75_6.0.out'
   !END

   !ATOM NAME='C_1   ' R=   -3.0480   -6.1140   -1.6410   !END
   !ATOM NAME='S_2   ' R=   -1.6620   -5.3230   -2.6230   !END
   !ATOM NAME='C_3   ' R=    6.1520    0.1110   -1.9800   !END
   !ATOM NAME='O_4   ' R=    6.9900   -0.7250   -2.3360   !END
   !ATOM NAME='N_5   ' R=    4.9020    0.1340   -2.4280   !END
   !ATOM NAME='C_6   ' R=    4.3330   -0.9460   -3.2100   !END
   !ATOM NAME='C_7   ' R=    2.9730   -0.5410   -3.7480   !END
   !ATOM NAME='S_8   ' R=    1.9100    0.0410   -2.6190   !END
   !ATOM NAME='C_9   ' R=   -4.1690    1.8230   -0.6330   !END
   !ATOM NAME='S_10  ' R=   -4.0350    0.1890   -0.0870   !END
   !ATOM NAME='FE11  ' R=   -0.2150   -2.2030    0.4530   !END
   !ATOM NAME='FE12  ' R=   -1.9980   -1.0120   -0.6880   !END
```

```
52      !ATOM NAME='FE13  ' R=    -1.1120    -3.3420    -1.8280  !END
53      !ATOM NAME='FE14  ' R=     0.2510    -1.2020    -1.8560  !END
54      !ATOM NAME='S_15  ' R=     0.0000     0.0000     0.0000  !END
55      !ATOM NAME='S_16  ' R=    -2.1570    -3.0770     0.1280  !END
56      !ATOM NAME='S_17  ' R=    -1.6670    -1.3870    -2.8250  !END
57      !ATOM NAME='S_18  ' R=     1.0880    -3.1400    -1.1970  !END
58      !ATOM NAME='FE19  ' R=     2.3480     0.4870    -0.3730  !END
59      !ATOM NAME='FE20  ' R=     1.3660     2.2890     1.1810  !END
60      !ATOM NAME='FE21  ' R=     3.7320     1.3360     1.6990  !END
61      !ATOM NAME='FE22  ' R=     1.5890    -0.1450     1.8530  !END
62      !ATOM NAME='S_23  ' R=     3.2450     2.4740    -0.0660  !END
63      !ATOM NAME='S_24  ' R=     2.0430     1.6490     3.1810  !END
64      !ATOM NAME='S_25  ' R=     3.4970    -0.8660     0.8140  !END
65      !ATOM NAME='C_26  ' R=     6.7200     3.1510     2.2200  !END
66      !ATOM NAME='S_27  ' R=     5.9030     1.6220     2.5740  !END
67      !ATOM NAME='C_28  ' R=     0.6150    -5.9900     1.5530  !END
68      !ATOM NAME='O_29  ' R=     1.6630    -6.5580     1.2240  !END
69      !ATOM NAME='N_30  ' R=     0.5620    -4.9000     2.3140  !END
70      !ATOM NAME='C_31  ' R=     1.7400    -4.2250     2.8260  !END
71      !ATOM NAME='C_32  ' R=     1.2890    -3.0080     3.6340  !END
72      !ATOM NAME='S_33  ' R=     0.2240    -1.8580     2.5740  !END
73      !ATOM NAME='C_34  ' R=    -0.8370     4.5870     2.2170  !END
74      !ATOM NAME='S_35  ' R=    -0.6080     3.4460     1.1620  !END
75      !ATOM NAME='C_36  ' R=     3.6140     4.8660     2.4130  !END
76      !ATOM NAME='O_37  ' R=     2.1640     4.8670     1.9430  !END
77      !ATOM NAME='H_38  ' R=    -1.4240     4.1735     3.1024  !END
78      !ATOM NAME='H_39  ' R=    -1.4240     5.4374     1.7356  !END
79      !ATOM NAME='H_40  ' R=     0.1709     4.9780     2.5786  !END
80      !ATOM NAME='H_41  ' R=    -5.1122     1.9370    -1.2629  !END
81      !ATOM NAME='H_42  ' R=    -3.2568     2.0891    -1.2629  !END
82      !ATOM NAME='H_43  ' R=    -4.2262     2.5207     0.2666  !END
83      !ATOM NAME='H_44  ' R=     4.2621     4.3283     1.6446  !END
84      !ATOM NAME='H_45  ' R=     3.6881     4.3283     3.4155  !END
85      !ATOM NAME='H_46  ' R=     3.9761     5.9405     2.5303  !END
86      !ATOM NAME='H_47  ' R=     6.0238     3.8232     1.6174  !END
87      !ATOM NAME='H_48  ' R=     7.6657     2.9458     1.6174  !END
88      !ATOM NAME='H_49  ' R=     6.9968     3.6690     3.1970  !END
89      !ATOM NAME='H_50  ' R=    -2.6058    -6.7615    -0.8135  !END
90      !ATOM NAME='H_51  ' R=    -3.6820    -6.7615    -2.3324  !END
91      !ATOM NAME='H_52  ' R=    -3.6992    -5.3000    -1.1795  !END
92      !ATOM NAME='H_53  ' R=     2.2063    -2.4391     4.0008  !END
93      !ATOM NAME='H_54  ' R=     0.6768    -3.3533     4.5315  !END
94      !ATOM NAME='H_55  ' R=     2.3954    -3.8862     1.9569  !END
95      !ATOM NAME='H_56  ' R=     2.3272    -4.9389     3.4930  !END
96      !ATOM NAME='H_57  ' R=    -0.4202    -4.4615     2.5878  !END
97      !ATOM NAME='H_58  ' R=    -0.3893    -6.3979     1.2003  !END
98      !ATOM NAME='H_59  ' R=     2.5011    -1.4462    -4.2553  !END
99      !ATOM NAME='H_60  ' R=     3.1018     0.2824    -4.5257  !END
100     !ATOM NAME='H_61  ' R=     4.2192    -1.8698    -2.5519  !END
101     !ATOM NAME='H_62  ' R=     5.0240    -1.1872    -4.0840  !END
102     !ATOM NAME='H_63  ' R=     4.2491     1.0041    -2.2070  !END
103     !ATOM NAME='H_64  ' R=     6.4367     0.9187    -1.2276  !END
104     !ATOM NAME='H_65  ' R=     1.8126     5.9137     1.8291  !END
105
106     !ISOLATE NF=3 RC=0.5 RCFAC=1.5 GMAX2=3.0 DECOUPLE=T !END
107
108     !CONSTRAINTS
109         !FREEZE ATOM='C_1'  !END
110         !FREEZE ATOM='C_7'  !END
111         !FREEZE ATOM='C_9'  !END
112         !FREEZE ATOM='C_26'  !END
113         !FREEZE ATOM='C_32'  !END
114         !FREEZE ATOM='C_34'  !END
115         !FREEZE ATOM='C_36'  !END
116     !END
117
```

```
!END
!EOB
```

C. Amino acids

In this chapter a general overview about amino acids is presented. Proteins in biological systems are all built by sequences of twenty different essential amino acids, the standard amino acids.. Table C.1 shows these amino acids.

Standard amino acids belong to the group of α-amino acids. They have a primary amino group (-NH_2) and a carboxyl group (-COOH) as substituents at the carbon atom (the so-called α-carbon). Proline is the only exception among the standard amino acids, which has a secondary amino group (-NH-). The amino and carboxyl groups of amino acids ionize easily. At physiological pH-values the amino acids are zwitterions. Figure C.1 shows such a zwitterion. Amino acids can build chains. A water molecule is eliminated and a peptide

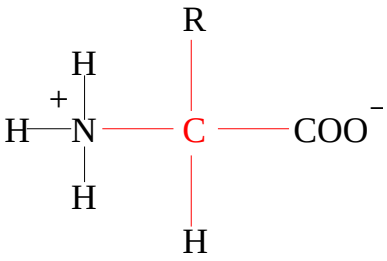


Figure C.1: The zwitterionic form of an amino acid.

bond is formed. The condensation reaction is shown in Figure C.2. Peptide

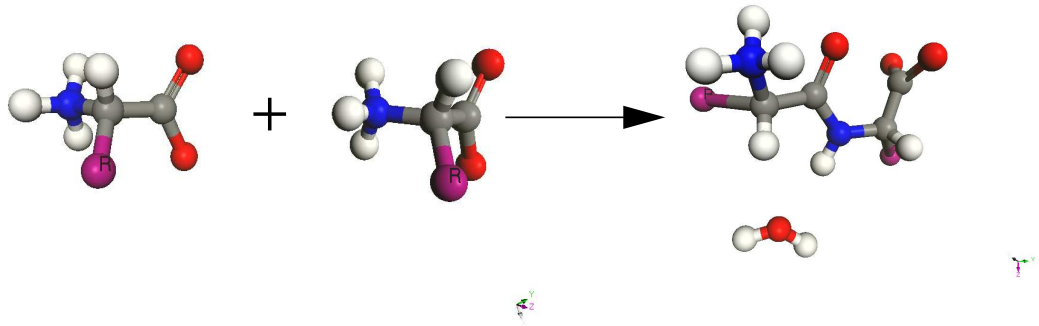
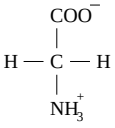
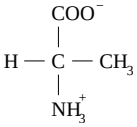
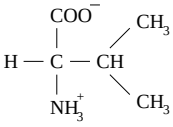
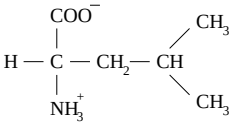
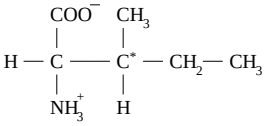
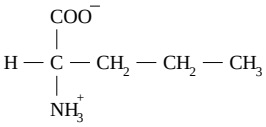
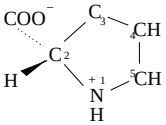
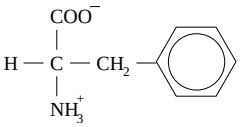
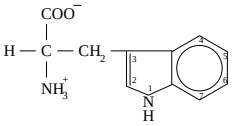


Figure C.2: The formation of a peptide bond by elimination of a water molecule from two amino acids.

bonds are characterized by a CO-NH bond. Small chains of amino acids (up to ten amino acids) are call di-, tri-, oligo- and polypeptides. Peptides are linear polymers. Each amino acid is linked to two other amino acids. This defines a

'head' and a 'tail' of a peptide chain. By definition the N-terminus is the head which is always at the left end of the peptide. The C-terminus is at the right end. Proteins are peptides with at least 100 amino acids. Variation in the length and in the sequence of peptides are the reason for the forms and functions of proteins.

Name	Code	Structure
non-polar side chains		
Glycine	GLY	
Alanine	ALA	
Valine	VAL	
Leucine	LEU	
Isoleucine	ILE	
Methionine	MET	
Proline	PRO	
Phenylalanine	PHE	
Tryptophane	TRP	

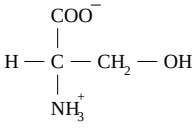
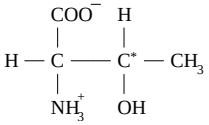
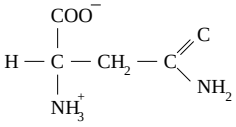
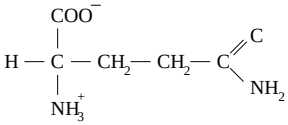
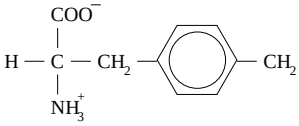
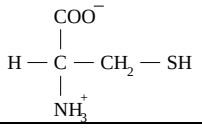
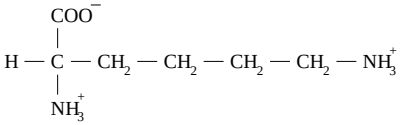
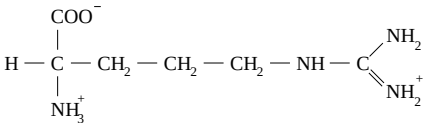
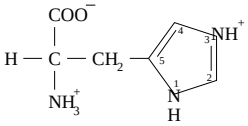
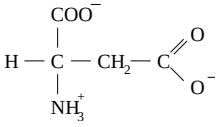
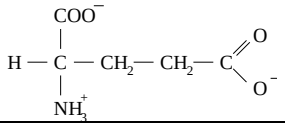
Name	Code	Structure
uncharged polar side chains		
Serine	SER	
Threonine	THR	
Asparagine	ASN	
Glutamine	GLN	
Tyrosine	TYR	
Cysteine	CYS	
charged polar side chains		
Lysine	LYS	
Arginine	ARG	
Histidine	HIS	
Asparagine acid	ASP	
Glutamine acid	GLU	

Table C.1: Twenty essential amino acids.

D. File Formats

D.1. The structure of PDB Files

For QMMM calculations CP-PAW needs to know which atoms belong to which subsystem. It is important to distinguish between MM system, QM system and link atoms. Therefore I expanded the PDB file format with some new columns to provide any information needed. Table D.1 describes the new format. The FORTRAN format describer is given by:

```
PDB_FORM= '(A6,I5,1X,A5,A4,A1,I4,4X,3F8.3, &
&           2F6.2,6X,A4,A2,2X,A1,1X,A12,A7)'
```

Column	Data Type	Description
1-6	character(6)	ATOM or HETATM. ATOM belongs to amino acids, different amino acids become connected via peptide bonds. HETATM are all other atoms in the system (e.g. water).
7-11	integer	atom serial number
13-17	character(5)	atom name (attention: this is not neccessarily the element name!)
18-21	character(4)	name of the residue
22	character(1)	identifer of the protein chain
23-26	integer	residue sequence number
31-54	real(8)	xyz-coordinates in angstrom
55-60	real(8)	occupancy (I don't use this.)
61-66	real(8)	temperature factor (I don't use this.)
73-76	character(4)	segment identifier (I don't use this.)
77-78	character(2)	Element symbol. Not all PDB files use this column.
81	character(1)	FLAG. This can be set to 'F' for fixed atoms, 'Q' for QM atoms or 'L' for link atoms.
83-94	character(12)	if flag is set Q or L then you can find the corresponding PAW name of the atom in this column. (The same name as in the structure file.)
96-102	character(7)	name of the link atom. Something like 'H_d12'

Table D.1: Expanded PDB file format for QMMM calculations.

E. Iron-iron distances of the model systems

	FE 1 FE11	FE 2 FE12	FE 3 FE13	FE 4 FE14	FE 5 FE19	FE 6 FE20	FE 7 FE21	FE 8 FE22	S_1 S_11
FE11		2.429	2.703	2.559	3.806	4.817	5.446	3.074	2.259
T1		0.163	0.034	-0.015	0.079	-0.249	0.070	-0.111	0.160
T2		0.331	0.017	0.012	-0.028	-0.311	-0.137	-0.313	0.214
T3		0.343	-0.087	-0.072	0.070	1.071	-0.062	-0.091	0.131
T4		0.331	0.065	0.105	-0.118	-0.328	-0.241	-0.415	0.074
T5		0.399	-0.004	0.100	-0.108	-0.410	-0.424	-0.629	0.199
T6		0.263	-0.004	0.065	0.035	-0.366	-0.146	-0.328	0.115
FE12			2.741	2.541	4.608	5.070	6.637	4.481	2.343
T1			-0.015	0.041	-0.042	-0.236	0.067	0.108	0.145
T2			-0.084	0.206	-0.080	-0.377	-0.104	0.060	0.100
T3			0.009	0.229	0.043	1.069	0.015	0.176	0.152
T4			-0.049	0.274	-0.202	-0.360	-0.186	0.007	0.114
T5			-0.048	0.222	-0.128	-0.384	-0.191	-0.080	0.107
T6			-0.043	0.154	-0.017	-0.197	0.006	0.085	0.109
FE13				2.537	5.362	6.849	7.602	5.574	3.968
T1				0.198	0.148	-0.145	0.167	-0.044	0.266
T2				0.183	-0.044	-0.340	-0.052	-0.257	0.143
T3				0.076	0.056	1.076	-0.074	-0.149	0.064
T4				0.166	-0.345	-0.390	-0.292	-0.371	0.065
T5				0.231	-0.144	-0.405	-0.285	-0.560	0.059
T6				0.252	-0.042	-0.281	-0.110	-0.317	0.106
FE14					3.074	4.760	5.585	4.082	2.225
T1					-0.125	-0.180	-0.074	-0.171	0.199
T2					-0.312	-0.235	-0.271	-0.305	0.252
T3					-0.098	1.130	-0.211	-0.206	0.166
T4					-0.630	-0.346	-0.569	-0.401	0.231
T5					-0.407	-0.270	-0.375	-0.403	0.109
T6					-0.365	-0.294	-0.336	-0.361	0.111
FE19						2.574	2.632	2.435	2.427
T1						0.004	0.093	0.104	-0.012
T2						0.167	0.092	0.151	0.039
T3						0.908	-0.051	0.035	-0.012
T4						0.246	0.070	0.221	0.025
T5						0.184	0.139	0.229	-0.102
T6						0.143	0.107	0.179	0.087
FE20							2.603	2.535	2.916
T1							0.108	0.026	-0.424
T2							0.066	0.186	-0.503
T3							0.069	0.935	0.996
T4							0.085	0.226	-0.460

	FE 1 FE11	FE 2 FE12	FE 3 FE13	FE 4 FE14	FE 5 FE19	FE 6 FE20	FE 7 FE21	FE 8 FE22	S_1 S_11
T5							0.089	0.273	-0.469
T6							0.048	0.049	-0.371
FE21								2.610	4.313
T1								0.123	-0.077
T2								0.105	-0.207
T3								-0.019	-0.151
T4								0.153	-0.287
T5								0.094	-0.289
T6								0.092	-0.093
FE22									2.445
T1									-0.016
T2									0.024
T3									-0.029
T4									-0.115
T5									-0.002
T6									0.023

Table E.1: The table shows the deviation of the Fe-Fe distances compared to the 3MIN structure.

List of Figures

Bibliography

[1] S. L. Miller and H. C. Urey: Organic Compound Synthes on the Primitive Earth: Several questions about the origin of life have been answered, but much remains to be studied. *Science* **130**(3370), 245–251 (1959).

[2] Johannes Kästner: *Biological Nitrogen Fixation - Simulation of the Reaction Mechanism of Nitrogenase from First Principles* PhD thesis Clausthal, University of Technology (2004).

[3] A. Warshel and M. Levitt: Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**, 227 (1976).

[4] U. C. Singh and P. A. Kollman: A Combined *Ab Initio* Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $CH_3Cl + Cl^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comp. Chem.* **7**(6), 718–730 (1986).

[5] B. Waszkowycz, I. H. Hillier, N. Gensmantel, and D. W. Payling: Combined Quantum Mechanical-Molecular Mechanical Study of Catalysis by the Enzyme Phospholipase A_2: An Investigation of the Potential Energy Surface for Amide Hydrolysis. *J. Chem. Soc. Perkin Trans.* **2**, 2025–2032 (1991).

[6] R. V. Stanton, D. S. Hartsough, and Jr. K. M. Merz: Calculation of Solvation Free Energies Using a Density Functional/Molecular Dynamics Coupled Potential. *J. Phys. Chem.* **97**, 11868–11870 (1993).

[7] M. J. Field, P. A. Bash, and M. Karplus: A Combined Quantum Mechanical and Molecular Mechnanical Potential for Molecular Dynamics Simulations. *J. Comp. Chem.* **11**(6), 700–733 (1990).

[8] Y. S. Lee, M. Hodoscek, B. R. Brooks, and P. F. Kador: Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics. *Biophys. Chem.* **70**, 203–216 (1998).

[9] P. D. Lyne, M. Hodoscek, and M. Karplus: A Hybrid QM-MM Potential Employing Hartree-Fock or Density Functional Methods in the Quantum Region. *J. Phys. Chem. A* **103**, 3462–3471 (1999).

[10] A. J. Turner, V. Moliner, and I. H. Williams: Transition-state structural refinement with GRACE and CHARMM: Flexible QM/MM modelling for lactate dehydrogenase. *Phys. Chem. Chem. Phys.* **1**, 1323–1331 (1999).

[11] D. Bakowies and W. Thiel: Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **100**, 10580–10594 (1996).

[12] J. H. Harding, A. H. Harker, P. B. Keegstra, R. Pandey, J. M. Vail, and C. Woodward: Hartree-Fock Cluster Computations of Defect and Perfect Ionic Crystal Properties. *Physica* **131**(B), 151–156 (1985).

[13] Z. Barandiaran and L. Seijo: The *ab initio* model potential representation of the crystalline environment. Theoretical study of the local distortion on NaCl:Cu^+. *J. Chem. Phys.* **89**(9), 5739–5746 (1988).

[14] D. Wei and D. R. Salahub: A combined functional and molecular dynamics simulation of a quantum water molecule in aqueous solution. *Chem. Phys. Let.* **224**, 291–296 (1994).

[15] Uwe Eichler, Christoph M. Kölmel, and Joachim Sauer: Combining ab initio techniques with analytical potential functions for structure predictions of large systems: Method and application to crystalline silica polymorphs. *Journal of Computational Chemistry* **18**(4), 463–477 (1997).

[16] X. P. Long, J. B. Nicholas, M. F. Guest, and R. L. Ornstein: A combined density functional theory/molecular mechanical formalism and its application to small water clusters. *J. Mol. Struc.* **412**, 121–133 (1997).

[17] T. K. Woo, P. M. Margl, P. E. Blöchl, and T. Ziegler: A Combined Car-Parrinello QM/MM implementation for *ab-initio* molecular dynamics simulations of extended systems: Application to transition metal catalysis. *J. Phys. Chem.* **101**(40), 7877–7880 (1997).

[18] J. Hutter M. Eichinger, P. Tavan and M. Parrinello: A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* **110**, 10452 (1999).

[19] T. K. Woo, P. E. Blöchl, and T. Ziegler: Towards solvation simulations with a combined ab initio molecular dynamics and molecular mechanics approach. *J. Theo. Chem.* **506**, 313–334 (2000).

[20] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus: CHARMM - A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**(3), 187–217 (1983).

[21] C. Bayly I. Gould K. Merz Jr. D. Ferguson D. Spellmeyer T. Fox J. Caldwell P. Kollman W. Cornell, P.Cieplak: A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

[22] J. Herman, H. J. C. Berendsen, W. F. v. Gusteren, and J. P. M. Postma: A Consistent Empirical Potential for Water-Protein Interactions. *Biopolymers* **23**, 1513–1518 (1984).

[23] J. H. Lii and N. L. Allinger: The MM3 force-field for amides, polypeptides and proteins. *J. Comp. Chem.* **12**(2), 186–199 (1991).

[24] M. Sierka and J. Sauer: Structure and reactivity of silica and zeolite catalysts by a combined quantum mechanics-shell-model potential approach based on DFT. *Faraday. Discuss.* **106**, 41–62 (1997).

[25] H. M. Senn and W. Thiel: QM/MM Methods for Biological Systems. *Top. Curr. Chem.* **268**, 173–290 (2007).

[26] M. J. S. Dewar and W. Thiel: Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **99**, 4899–4907 (1977).

[27] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart: Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902 (1985).

[28] ChemShell: a Computational Chemistry Shell see chemshell.org.

[29] I. Antes and W. Thiel: Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods. *J. Phys. Chem. A* **103**(46), 9290–9295 (1999).

[30] Y. K. Zhang, T. S. Lee, and W. T. Yang: A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **110**(1), 46–54 (1999).

[31] D. M. Philipp and R. A. Friesner: Mixed *ab initio* QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comp. Chem.* **20**(14), 1486–1494 (1999).

[32] K. P. Eurenius, D. C. Chatfield, B. R. Brooks, and M. Hodoscel: Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *Int. J. Quant. Chem.* **60**(6), 1189–1200 (1996).

[33] F. Maseras and K. Morokuma: IMOMM: A New Integrated Ab Initio + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comp. Chem.* **16**, 1170–1179 (1995).

[34] T. K. Woo, L. Cavallo, and T. Ziegler: Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamic simulations and frequency calculations. *Theor. Chem. Acc.* **100**, 307–313 (1998).

[35] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, and G. J. Kramer: Zeolite structure and reactivity by combined quantum-chemical-classical calculations. *J. Phys. Chem. B* **103**(29), 6133–6141 (1999).

[36] N. L. Allinger, M. T. Tribble, M. A. Miller, and D. H. Wertz: Conformational Analysis. LXIX. An Improved Force Field for the Calculation of the Structures and Energies of Hydrocarbons. *J. Am. Chem. Soc.* **93**(7), 1637–1648 (1971).

[37] L. S. Bartell: Representations of Molecular Force Fields. 3. On Gauche Conformational Energy. *J. Am. Chem. Soc.* **99**(10), 3279–3282 (1977).

[38] N. L. Allinger, Y. H. Yuh, and J.-H. Lii: Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **111**(23), 8551–8566 (1989).

[39] J.-H. Lii and N. L. Allinger: Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 2. Vibrational Frequencies and Thermodynamics. *J. Am. Chem. Soc.* **111**(23), 8566–8575 (1989).

[40] J.-H. Lii and N. L. Allinger: Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 3. The van der Waals' Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons. *J. Am. Chem. Soc.* **111**(23), 8576–8582 (1989).

[41] N. L. Allinger, K. Chen, and J.-H. Lii: An Improved Force Field (MM4) for Saturated Hydrocarbons. *J. Comput. Chem.* **17**(5-6), 642–668 (1996).

[42] N. Nevins, K. Chen, and N. L. Allinger: Molecular Mechanics (MM4) Calculations on Alkenes. *J. Comput. Chem.* **17**(5-6), 669–694 (1996).

[43] N. Nevins, J.-H. Lii, and N. L. Allinger: Molecular Mechanics (MM4) Calculations on conjugated Hydrocarbons. *J. Comput. Chem.* **17**(5-6), 695–729 (1996).

[44] N. Nevins and N. L. Allinger: Molecular Mechanics (MM4) Vibrational Frequency Calculations for Alkenes and Conjugated Hydrocarbons. *J. Comput. Chem.* **17**(5-6), 730–746 (1996).

[45] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner: A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**(3), 765–784 (1984).

[46] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives: Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).

[47] T. A. Halgren: Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. *J. Comp. Chem.* **17**(5-6), 490–519 (1996).

[48] T. A. Halgren: Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comp. Chem.* **17**(5-6), 520–552 (1996).

[49] T. A. Halgren: Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comp. Chem.* **17**(5-6), 553–586 (1996).

[50] T. A. Halgren and R. B. Nachbar: Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comp. Chem.* **17**(5-6), 587–615 (1996).

[51] T. A. Halgren: Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data and Emperical Rules. *J. Comp. Chem.* **17**(5-6), 616–641 (1996).

[52] T. A. Halgren: MMFF VII. Characterization of MMFF94, MMFF94s and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interactio Energies and Geometries. *J. Comp. Chem.* **20**(7), 730–748 (1999).

[53] Villoya S. Allured, Christine M. Kelly, and Clark R. Landis: SHAPES Empirical Force Field: New Treatment of Angular Potentials and Its Application to Square-Planar Transition-Metal Complexes. *J. Am. Chem. Soc.* (1991).

[54] A. J. Stone: *The Theory of Intermolecular Forces* Oxford Press (1996).

[55] T. L. Hill: Steric Effects. I. Van der Waals Potential Energy Curves. *J. Chem. Phys.* **16**(4), 399–405 (1948).

[56] D. Hall and N. Pavitt: An Appraisal of Molecular Force Fields for the Representation of Polypeptides. *J. Comp. Chem.* **5**(5), 441–450 (1984).

[57] L. Nilsson and M. Karplus: Empirical Energy Functions for Energy Minimization and Dynamics of Nucleic Acids. *J. Comp. Chem.* **7**(5), 591–616 (1986).

[58] W. L. Jorgensen and J. Tirado-Rives: The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **110**(6), 1657–1666 (1988).

[59] J. Tirado-Rives and W. L. Jorgensen: Molecular Dynamics of Proteins with the OPLS Potential Functions. Simulation of the Third Domain of Silver Pheasant Ovomucoid in Water. *J. Am. Chem. Soc.* **112**, 2773–2781 (1990).

[60] M. Orozco, J. Tirando-Rives, and W. L. Jorgensen: Mechanism for the Rotamase Activity of FK506 Binding Protein from Molecular Dynamics Simulations. *Biochemistry* **32**, 12864–12874 (1993).

[61] P. C. Hariharan and J. A. Pople: The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theoret. chim. Acta* **28**, 213–222 (1973).

[62] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein: Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**(2), 926–935 (1982).

[63] L. Kuyper, D. Ashton, K. M. Merz Jr., and P. A. Kollman: Free Energy Calculations on the Relative Solvation Free Energies of Benzene, Anisole, and 1,2,3-Trimethoxybenzene: Theoretical and Experimental Analysis of Aromatic Methoxy Solvation. *J. Phys. Chem.* **95**, 6661–6666 (1991).

[64] C. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman: A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

[65] W. Cornell, P. Cieplak, C. Bayly, and P. A. Kollman: Application of RESP Charges To Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **115**, 9620–9631 (1993).

[66] A. Hagler, E. Euler, and S. Lifson: Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals. *J. Am. Chem. Soc.* **96**, 5319–5327 (1974).

[67] I. R. Gould and P. A. Kollman: Ab Initio SCF and MP2 Calculations on Four Low-Energy Conformers of N-Acetyl-N'-methylalaninamide. *J. Phys. Chem.* **96**, 9255–9258 (1992).

[68] I. R. Gould, W. D. Cornell, and I. H. Hillier: A Quantum Mechanical Investigation of the Conformational Energetics of the Alanine and Glycine Dipeptides in the Gas Phase and in Aquaeous Solution. *J. Am. Chem. Soc.* **116**, 9250–9256 (1994).

[69] T. A. Halgren: Maximally Diagonal Force Constants in Dependent Angle-Bending Coordinate. 2. Implications for the Design of Empirical Force Fields. *J. Am. Chem. Soc.* **112**, 4710–4723 (1990).

[70] J. Caldwell and P. Kollman: The Structures and Properties of Neat Liquids using Nonadditive Molecular Dynamics: Water, Methanol and N-Methyl Acetamide. *J. Phys. Chem.* **99**, 6208–6219 (1995).

[71] W. L. Jorgensen: Transferable Intermolecular Potential Functions for Water, Alcohols and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **103**(2), 335–340 (1981).

[72] W. L. Jorgensen: Revised TIPS for simulations of liquid water and aqueous solutions. *J. Chem. Phys.* **77**(7), 4156–4163 (1982).

[73] A. T. Hagler, E. Huler, and S. Lifson: Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals. *J. Am. Chem. Soc.* **96**(17), 5319–5327 (1973).

[74] A. H. Narten and H. A. Levy: Liquid Water: Molecular Correlation Functions from X-Ray Diffraction. *J. Chem. Phys.* **55**(5), 2263–2269 (1971).

[75] H. Federmann: QM-MM Hybridsimulationen komplexer Systeme. Implementierung einer Wasserumgebung Master's thesis Clausthal, University of Technology (2007).

[76] P. Hohenberg and W. Kohn: Inhomogeneous Electron Gas. *Phys. Rev.* **136**(3B), 864–871 (1964).

[77] R. G. Parr and W. Yang: *Density-Functional Theory of Atoms and Molecules* Oxford University Press (1989).

[78] W. Kohn and L. J. Sham: Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**(4A), 1133–1138 (1965).

[79] S. K. Ghosh and R. G. Parr: Density-determined orthonormal orbital approach to atomic energy funtionals. *J. Chem. Phys.* **82**, 3307–3315 (1985).

[80] J. P. Perdew and Y. Wang: Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **45**(23), 13244–13249 (1992).

[81] A. D. Becke: Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **96**, 2155 (1992).

[82] A. D. Becke: Density-functional thermochemistry. II. The effect of the Perdew-Wang generalized-gradient correlation correction. *J. Chem. Phys.* **97**, 9173 (1992).

[83] J.P. Perdew, K. Burke, and M. Ernzerhof: Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

[84] P. E. Blöchl, C. J. Först, and J. Schimpl: Projector augmented wave method: *ab initio* molecular dynamics with full wave functions. *Bull. Mat. Sci.* **26**, 33 (2003).

[85] Peter Blöchl: Projector augmented-wave method. *Physical Review B* **50**(24), 17953–17979 (1994).

[86] L. Verlet: Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).

[87] R.Car and M. Parrinello: Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters* **55**, 2471–2474 (1985).

[88] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson: Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).

[89] J. M. Word, S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, and D. C. Richardson: Visualizing and Quantifying Molecular Goodnes-of-Fit: Small-probe Contact Dots with Explicit Hydrogen Atoms. *J. Mol. Biol.* **285**, 1711–1733 (1999).

[90] K. Morokuma and L. Pedersen: Molecular-Orbital Studies of Hydrogen Bonds. An *Ab Initio* Calculation for Dimeric H_2O. *J. Chem. Phys.* **48**(7), 3275–3282 (1968).

[91] A. D. Buckingham: The hydrogen bond, and the structure and properties of H_2O and $(H_2O)_2$. *J. Mol. Struc.* **250**, 111–118 (1991).

[92] J. A. Odutola and T. R. Dyke: Paritally deuterated water dimers: Microwave spectra and structure. *J. Chem. Phys.* **72**(9), 5062–5070 (1980).

[93] S. Scheiner: Ab Initio Studies of Hydrogen Bonds: The Water Dimer Paradigm. *Annu. Rev. Phys. Chem.* **45**, 23–56 (1994).

[94] J. Talafous, K. J. Marcinowski, G. Klopman, and M. G. Zagorski: Solution Structure of Residues 1-28 of th Amyloid β-Peptide. *Biochemistry* **33**, 7788–7796 (1994).

[95] R.N.F. Thorneley and D.J. Lowe: The mechanism of *Klebsiella pneumoniae* nitrogenase action. Pre-steady-state kinetics of an enzyme-bound intermediate in N_2 reduction and of NH_3 formation. *Biochem. J.* **224**, 887 (1984).

[96] R.N.F. Thorneley and D.J. Lowe: The mechanism of *Klebsiella pneumoniae* nitrogenase action. Simulation of the dependences of H_2 evolution rate on component-protein concentration and ratio and sodium dithionite concentration. *Biochem. J.* **224**, 903 (1984).

[97] D.J. Lowe and R.N.F. Thorneley: The mechanism of *Klebsiella pneumoniae* nitrogenase action. The determination of rate constants required for the simulation of the kinetics of N_2 reduction and H_2 evolution. *Biochem. J.* **224**, 895 (1984).

[98] D.J. Lowe and R.N.F. Thorneley: The mechanism of Klebsiella pneumoniae nitrogenase action. Pre-steady-state kinetics of H2 formation. *Biochem. J.* **224**, 877 (1984).

[99] R.A. Alberty: Thermodynamiks of the Nitrogenase Reaction. *J. Biol. Chem.* **269**, 7099 (1994).

[100] I. V. Kurnikov, A. K. Charnley, and D. N. Beratan: From ATP to Electron Transfer: Electrostatics and Free-Energy Transduction in Nitrogenase. *J. Phys. Chem. B* **105**, 5359 (2001).

[101] D. Voet and J.G. Voet: *Biochemistry* John Wiley and Sons (1990).

[102] J. Schimpl, H.M. Petrilli, and P.E. Blöchl: Nitrogen Binding to the FeMo-Cofactor of Nitrogenase. *J. Am. Chem. Soc.* **125**, 15772 (2003).

[103] F.B. Simpson and R.H. Burris: A Nitrogen Pressure of 50 Atomspheres Does Not Prevent Evolution of Hydrogen by Nitrogenase. *Science* **224**, 1095 (1984).

[104] J.M. Rivera-Ortiz and R.H. Burris: Ineractions Among Substrates and Inhibitors of Nitrogense. *J. Bacteriol.* **123**, 537 (1975).

[105] D.C. Rees and J.B. Howard: Nitrogenase: Standing at the crossroads. *Curr. Opin. Chem. Biol.* **4**, 559 (2000).

[106] J.B. Howard and D.C. Rees: Structural Basis of Biological Nitrogen Fixation. *Chem. Rev.* **96**, 2965 (1996).

[107] J.H. Spee, A.F. Arendsen, H. Wassink, S.J. Marrit, W.R. Hagen, and H. Haaker: Redox Properties and electron paramagnetic resonance spectroscopy of the transition state complex of Azobacter vinelandii nitrogenase. *FEBS Letters* **432**, 55 (1998).

[108] B. Schmid, O. Einsle, H. J. Chiu, A. Willing, M. Yoshida, J. B. Howard, and D. C. Rees: Biochemical and Structural Characterization of the Crosslinked Complex of Nitrogenase: Comparison to the ADP-ALF4-Stabilized Structure. *Biochemistry* **41**, 15557 (2002).

[109] O. Einsle, F.A. Tezcan, S.L.A. Andrade, B. Schmid, M. Yoshida, J.B. Howard, and D.C. Rees: Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science* **297**, 1696 (2002).

[110] J. Kim and D.C. Rees: Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. *Nature* **360**, 553 (1992).

[111] B. Hinnemann and J.K. Nørskov: Modeling a Central Ligand in the Nitrogenase FeMo Cofactor. *J. Am. Chem. Soc.* **125**, 1466 (2003).

[112] I. Dance: The consequences of an interstitial N atom in the FeMo cofactor of nitrogenase. *Chem. Commun.* **3**, 324 (2003).

[113] H.I. Lee, P.M.C. Benton, M. Laryukhin, R.Y. Igarashi, D.R. Dean, L.C. Seefeldt, and B.M. Hoffman: The Interstitial Atom of the Nitrogenase FeMo-Cofactor: ENDOR and ESEEM Show It Is Not an Exchangeable Nitrogen. *J. Am. Chem. Soc.* **125**, 5604 (2003).

[114] T. Lovell, T. Liu, D.A. Case, and L. Noodleman: Structural, Spectroscopic, and Redox Consequences of a Central Ligand in the FeMoco of Nitrogenase: A Density Functional Theoretical Study. *J. Am. Chem. Soc.* **125**, 8377 (2003).

[115] B. Hinnemann and J.K. Nørskov: Chemical Activity of the Nitrogenase FeMo Cofactor with a Central Nitrogen Ligand: Density Functional Study. *J. Am. Chem. Soc.* **126**, 3920 (2004).

[116] T.H. Rod, B. Hammer, and J.K. Nørskov: Nitrogen Adsorption and Hydrogenation on a $MoFe_6S_9$ Complex. *Phys. Rev. Lett.* **82**, 4054 (1999).

[117] T.H. Rod and J.K. Nørskov: Modeling the Nitrogenase FeMo Cofactor. *J. Am. Chem. Soc.* **122**, 12751–12763 (2000).

[118] U. Huniar, R. Ahlrichs, and D. Coucouvanis: Density Functional Theory Calculations and Exploration of a Possible Mechanism of N_2 Reduction by Nitrogenase. *J. Am. Chem. Soc.* **126**, 2588 (2004).

[119] Johannes Kästner and Peter E. Blöchl: Model for Acetylene Reduction by Nitrogenase Derived from Density Functional Theory. *J. Inorg. Chem.* **44**, 4568–4575 (2005).

[120] Johannes Kästner, Sascha Hemmen, and Peter E. Blöchl: Activation and protonation of dinitrogen at the FeMo cofactor of nitrogenase. *J. Chem. Phys.* **123**, 074306 (2005).

[121] D. Sellmann and J. Sutter: Elementary reactions, structure-function relationships, and the potential relevance of low molecular metal-sulfur ligand complexes to biological N_2 fixation. *J. Biol. Inorg. Chem. (JBIC)* **1**, 597 (1996).

[122] D. Sellmann and J. Sutter: In Quest of Competitive Catalysts for Nitrogenases and Other Metal Sulfur Enzymes. *Acc. Chem. Res.* **30**, 460 (1997).

[123] D. Sellmann, J. Utz, N. Blum, and F.H. Heinemann: On the function of nitrogenase FeMo cofactors and competitive catalysts: chemical principles, structural blue-prints, and the relevance of iron sulfur complexes for N_2 fixation. *Coord. Chem. Rev.* **190-192**, 607 (1999).

[124] D. Sellmann, A. Fürsattel, and J. Sutter: The nitrogenase catalyzed N_2 dependent HD formation: a model reaction and its significance for the FeMoco function. *Coord. Chem. Rev.* **200-202**, 545 (2000).

[125] R.K. Szilagyi, D.G. Musaev, and K. Morokuma: Theoretical Studies of Biological Nitrogen Fixation. I. Density Functional Modeling of the Mo-Site of the FeMo-Cofactor. *Inorg. Chem.* **40**, 766 (2001).

[126] M.C. Durrant: A molybdenium-centered model for nitrogenase catalysis. *Inorg. Chem. Comm.* **4**, 60–62 (2001).

[127] M.C. Durrant: Controlled protonation of iron-molybdenum cofactor by nitrogenase: a structural and theoretical analysis. *Biochem. J.* **355**, 569 (2001).

[128] M.C. Durrant: An Atomic-Level Mechanism for Molybdenum Nitrogenase. Part 1. Reduction of Dinitrogen. *Biochemistry* **41**, 13934 (2002).

[129] M.C. Durrant: An Atomic-Level Mechanism for Molybdenum Nitrogenase. Part 2. Proton Reduction, Inhibition of Dinitrogen Reduction by Dihydrogen, and the HD Formation Reaction. *Biochemistry* **41**, 13946 (2002).

[130] D.V. Yandulov and R.R. Schrock: Catalytic Reduction of Dinitrogen to Ammonia at a Single Molybdenum Center. *Science* **301**, 76–78 (2003).

[131] J.W. Peters, M. Stowell, S.M. Soltis, M.G. Finnegan, M.K. Johnson, and D. Rees: Redox-Dependent Structural Changes in the Nitrogenase P-Cluster. *Biochem.* **36**, 1181–1187 (1997).

[132] S.M. Mayer, D.M. Lawson, C.A. Gormal, S.M. Roe, and B.E. Smith: New insights into structure-function relationships in nitrogenase: a 1.6 Å resolution X-ray crystallographic study of *Klebsiella pneumoniae* MoFe-protein. *J. Mol. Biol.* **292**, 871 (1999).

[133] K.K. Surerus, M.P. Hendrich, P. Christie, D. Rottgardt, W.H. Orme-Johnson, and E. Münck: Mössbauer and Integer Spin EPR of the Oxidized P-Clusters of Nitrogenase: P^{ox} is a Non-Kramers System with a Nearly Degenerate Ground Doublet. *J. Am. Chem. Soc.* **114**, 8579 (1992).

[134] S.J. Yoo, H.C. Angove, V. Papaefthymiou, B.K. Burgess, and E. Münck: Mössbauer Study of the MoFe Protein of Nitrogenase from *Azotobacter vinelandii* Using Selective ^{57}Fe Enrichment of the M-Centers. *J. Am. Chem. Soc.* **122**, 4926 (2000).

[135] J.-M. Mouesca, L. Noodleman, and D.A. Case: Analysis of the 57Fe Hyperfine Coupling Constants and Spin States in Nitrogenase P-Clusters. *Inorg. Chem.* **33**, 4819 (1994).

[136] M.K.Chan, J. Kim, and D.C. Rees: The Nitrogenase FeMo-Cofactor and P-Cluster Pair: 2.2 Å Resolution Structures. *Science* **260**, 792 (1993).

[137] R.C. Tittsworth and B.J. Hales: Detection of EPR Signals assigned to the 1-equiv-oxidized P-clusters of the nitrogenase MoFe protein from Azobacter vinelandii. *J. Am. Chem. Soc.* **115**, 9763 (1993).

[138] A.J. Pierik, H. Wassink, H. Haaker, and W.R. Hagen: Redox properties and EPR spectroscopy of the P-clusters of Azobacter-vinelandii. *J. Biochem.* **212**, 51 (1994).

[139] The protein data bank: www.pdb.org.

[140] Alexander Poddey: A dimer method for the determination of transition states Diplomarbeit Technische Universität Clausthal (2004).

[141] A. Klamt and G. Schüürmann: COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin. Trans.* **2**(5), 799–805 (1993).

[142] H. M. Senn, P. M. Margl, R. Schmid, T. Ziegler, and P. E. Blöchl: *Ab initio* molecular dynamics with a continuum solvation model. *J. Chem. Phys.* **118**(3), 1089–1100 (2003).

[143] P. E. Bloechl: Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *J. Chem. Phys.* **103**, 7422–7428 (1995).

[144] Persistence of Vision Raytracer POVRAY: www.povray.org.

[145] W. Humphrey, A. Dalke, and K. Schulten: VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **14**(1), 33–38 (1996).

[146] D. Frishman and P. Argos: Knowledge-based secondary structure assigment. *Proteins: structure, function and genetics* **23**, 566–579 (1995).

[147] OpenDX Data Explorer: www.opendx.org.

Acknowledgment

This thesis is the result of three year's hard work. During this time I have been supported by many people in different ways. Now it is time to express my gratitude to all of them.

First, I have to thank my advisor, Prof. Peter E. Blöchl. Whenever I needed advise he took the time to discuss the problems with me. His critical view often motivated me to think twice about a problem. Peter showed me that most problems can be solved even if they appear insolvable

I thank Johannes Kästner for introducing me into the world of Nitrogenase and force field simulations. He let me participate in his comprehensive knowledge of QM/MM simulations which has been a solid base to start my researches.

Special thanks goes to Alexander Poddey. During the time I have been part of the department he become a real friend. I will miss the time when we discussed problems about physics, computers (hard- and software) or just talked about nothing. From Alexander I have learned all about compiling and linking program code.

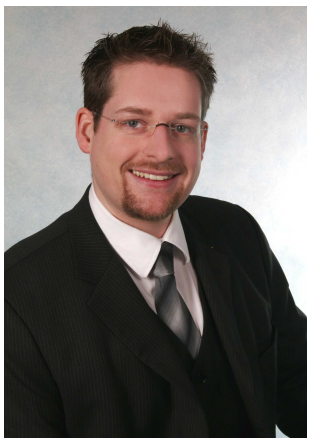
Many other people in the institute let me experience an atmosphere of being among friends. Jürgen Noffke, Christian Walther, Manuel Krämer and Holger Federmann are just a view who made my life in the institute more enjoyable. I thank Axel Ehrich and the other members of our IT team for their continuous support, especially in computer-related questions.

Clausthal, March 2008 Sascha Hemmen

Curriculum Vitae

Lebenslauf

Persönliche Daten



Name:	Sascha Hemmen
Anschrift:	Bohlweg 2a
	38678 Clausthal
Tel.(dienstl.):	+49 (0) 5323 - 72 2017
Tel.(privat):	+49 (0) 5323 - 718 475
Mobil:	+49 (0) 177 - 483 73 05
Email:	sascha.hemmen@tu-clausthal.de
Geburtsdatum, -ort:	26. Mai 1977, Saarbrücken
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

Berufspraxis

aktuell	Institut für Theoretische Physik, TU Clausthal Wissenschaftlicher Mitarbeiter • Promotion über die Funktion des P-Clusters im Enzym Nitrogenase sowie die Einbindung von Multiskalenmethoden in das CP-PAW Programmpaket • Lehrbetreuung • Diplomandenbetreuung • Organisation, Durchführung und Teilnahme an Workshops

Weiterbildung, Trainings, Seminare

10/2007	2. Harzer *Ab Initio* Workshop
10/2006–03/2007	Entscheidung und Organisation I
10/2006	1. Harzer *Ab Initio* Workshop
06/2006	CECAM PAW Workshop
10/2005–03/2006	Arbeitsrecht I & II
04/2005–09/2005	Qualitätsmanagment I
04/2004	Managementseminar

Studium

10/2004–03/2007	Promotionsstudiengang Physik Technische Universität Clausthal
10/1998–10/2004	Physik/Physikalische Technologien Technische Universität Clausthal Studienschwerpunkte: • Simulation und Modellierung • Theoretische Physik • Biochemie • Halbleiterphysik Diplomarbeit: Palladiumkatalysierte Arylierung heterocyclischer Alkene Abschluss: Diplom Physiker (Note: sehr gut)

Wehrdienst

11/1997–08/1998	Transportbataillon Munitions- und Betriebsstoffwart

Schulbildung

09/1988–06/1997	Gymnasium Edertalschule Frankenberg (Hessen) Abschluss: Abitur (Note 1,8)
09/1987–08/1988	Leibniz Gymnasium St. Ingbert (Saarland)
09/1983–08/1987	Südschule St. Ingbert (Saarland)

Sprachen

Englisch	sicher in Wort und Schrift (ständiger beruflicher Gebrauch, Sprache der Dissertation)

Fähigkeiten

Simulationssoftware, -techniken:	CP-PAW, QM-MM, COSMO: detaillierte Kenntnisse AMBER Kraftfeld: umfangreiche Kenntnisse GROMACS: Grundkenntnisse
Betriebssysteme:	Unix (Linux): umfangreiche Kenntnisse Windows: umfangreiche Kenntnisse
Office-Programme:	Microsoft Office Paket: umfangreiche Kenntnisse LaTeX: umfangreiche Kenntnisse
Programmiersprache	FORTRAN: umfangreiche Kenntnisse C, C++: Grundkenntnisse Java: Grundkenntnisse PHP, MySQL: Grundkenntnisse

Interessen

Mitglied im Arbeitskreis „Werbung" der Fakultät für Natur- und Materialwissenschaften der TU Clausthal zur Gewinnung neuer Studenten und Organisation des Sommerkollegs 2006 und 2007
Arbeit in einer studentischen Verbindung zur Weiterbildung der fachlichen und sozialen Kompetenz
Wandern, Kraft- und Ausdauertraining, Squash, Basketball
Weiterbildung durch Lesen von Fachliteratur

Sascha Hemmen